Results from evaluation of three commercial off-the-shelf face recognition systems on Chokepoint dataset

Prepared by:
E. Granger, D. Gorodnichy, E. Choy, W. Khreich, Pr. Radtke, J.-P. Bergeron, D. Bissessar

Canada Border Services Agency
Ottawa ON Canada K1A 0L8
and
Université du Québec
1100, rue Notre-Dame Ouest
Montréal (Québec) H3C 1K3

Scientific Authority: Pierre Meunier DRDC Centre for Security Science 613-992-0753

Contract Report
DRDC-RDDC-2014-C247
September 2014

IMPORTANT INFORMATIVE STATEMENTS

PROVE-IT (FRiV) Pilot and Research on Operational Video-based Evaluation of Infrastructure and Technology: Face Recognition in Video project, PSTP 03-401BIOM, was supported by the Canadian Safety and Security Program (CSSP) which is led by Defence Research and Development Canada's Centre for Security Science, in partnership with Public Safety Canada. Led by Canada Border Services Agency partners included: Royal Canadian Mounted Police, Defence Research Development Canada, Canadian Air Transport Security Authority, Transport Canada, Privy Council Office; US Federal Bureau of Investigation, National Institute of Standards and Technology, UK Home Office; University of Ottawa, Université Québec (ÉTS).

The CSSP is a federally-funded program to strengthen Canada's ability to anticipate, prevent/mitigate, prepare for, respond to, and recover from natural disasters, serious accidents, crime and terrorism through the convergence of science and technology with policy, operations and intelligence.

Science and Engineering Directorate

Border Technology Division

Division Report 2014-27 (TR)
July 2014

Results from evaluation of three commercial off-the-shelf face recognition systems on Chokepoint dataset

E. Granger, D. Gorodnichy, E. Choy, W. Khreich, Pr. Radtke, J.-P. Bergeron, D. Bissessar

This page left intentionally blank

Abstract

This report is a supplement to Division Report 2014-27 (TR) ' 'Evaluation methodology for face recognition technology in video surveillance applications" (by E. Granger and D. Gorodnichy) [1]. It presents complete evaluation results of three Commercial Off-The-Shelf (COTS) Face Recognition (FR) systems: Cognitec, PittPatt and Neurotechnology — obtained on the Chokepoint public data-set using the multi-level evaluation methodology introduced in the previous report. Four levels of details are examined according to the methodology for each target person from the Checkpoint data-set: Level 0 or score-based analysis illustrates the probability distribution of the genuine scores against that of the impostors at different face resolutions, which visually illustrates the discrimination power of the system for each target individual. Level 1 or transaction-based analysis provides the averaged description of the system in terms of false positive and false negative rates aggregated over all transactions, expressed using Receiver Operative Curves (ROC), Detection Error Trade-off (DET), and Precision-Recall Operating Characteristic (PROC) curves. Level 2 or subject-based analysis describes the performance of the system using the-so-called "Doddington's Zoo" categorization of individuals, which detects whether an individual belongs to an easier or a harder classes of people that the system is able to recognize [2, 3]. Finally, Level 3 or temporal analysis allows one to examine the overall discrimination power of the system by accumulating the positive predictions while tracking a person over time and computing the recognition confidence based thereon. Two-page Report Cards summarizing the performance of the system for each target individual are published, thus providing an exhaustive report of the systems performance on a variety of different target subjects. As highlighted in previous report [1] and other publications [4, 5], such exhaustive reporting of the biometric system performance is required when the variation of system performance from one target individual to another is suspected. Our results show that this is indeed the case for all three tested COTS FR systems.

Keywords: video-surveillance, face recognition in video, instant face recognition, watch-list screening, biometrics, reliability, performance evaluation

Community of Practice: Biometrics and Identity Management

Canada Safety and Security (CSSP) investment priorities:

1. Capability area: P1.6 – Border and critical infrastructure perimeter screening technologies/ protocols for rapidly detecting and identifying threats.
2. Specific Objectives: O1 – Enhance efficient and comprehensive screening of people and cargo (identify threats as early as possible) so as to improve the free flow of legitimate goods and travellers across borders, and to align/coordinate security systems for goods, cargo and baggage;
3. Cross-Cutting Objectives CO1 – Engage in rapid assessment, transition and deployment of innovative technologies for public safety and security practitioners to achieve specific objectives;
4. Threats/Hazards F – Major trans-border criminal activity – e.g. smuggling people/ material

Acknowledgements

This work is done within the project PSTP-03-401BIOM "PROVE-IT(FRiV)" funded by the Defence Research and Development Canada (DRDC) Centre for Security Science (CSS) Public Security Technical Program (PSTP) and in-kind contributions from École de technologie supérieure by the following contributors:

1. **J.-P. Bergeron, D. Bissessar, E. Choy, D. Gorodnichy**, Science & Engineering Directorate, Canada Border Services Agency,
2. **E. Granger, W. Khreich, P. Radtke,**, École de technologie supérieure, Université du Québec.

Disclaimer

In no way do the results presented in this paper imply recommendation or endorsement by the Canada Border Services Agency (CBSA), nor do they imply that the products and equipment identified are necessarily the best available for the purpose. The information presented in this report contains only the information available in public domain.

The results presented in this report were produced in experiments conducted by the CBSA, and should therefore not be construed as vendor's maximum-effort full-capability result. In no way do the results presented in this presentation imply recommendation or endorsement by the CBSA, nor do they imply that the products and equipment identified are necessarily the best available for the purpose. Additionally, it is noted that the Cognitec technology was operated under the conditions that are outside of its specifications.

Release Notes

Context: This document is part of the set of reports produced for the PROVE-IT(FRiV) project. All PROVE-IT(FRiV) project reports are listed below.

- Dmitry Gorodnichy and Eric Granger "PROVE-IT(FRiV): framework and results". Also published in Proceedings of NIST International Biometrics Performance Conference (IBPC 2014), Gaithersburg, MD, April 1-4, 2014. Online at http://www.nist.gov/itl/iad/ig/ibpc2014.cfm.
- Dmitry Gorodnichy and Eric Granger, "Evaluation of Face Recognition for Video Surveillance". Also published in Proceedings of NIST International Biometric Performance Conference (IBPC 2012), Gaithersburg, March 5-9, 2012. Online at http://www.nist.gov/itl/iad/ig/ibpc2012.cfm.
- D. Bissessar, E. Choy, D. Gorodnichy, T. Mungham, "Face Recognition and Event Detection in Video: An Overview of PROVE-IT Projects (BIOM401 and BTS402)", Border Technology Division, Division Report 2013-04 (TR).
- E. Granger, P. Radtke, and D. Gorodnichy, "Survey of academic research and prototypes for face recognition in video", Border Technology Division, Division Report 2014-25 (TR).
- D. Gorodnichy, E.Granger, and P. Radtke, "Survey of commercial technologies for face recognition in video", Border Technology Division, Division Report 2014-22 (TR).
- E. Granger and D. Gorodnichy, "Evaluation methodology for face recognition technology in video surveillance applications", Border Technology Division, Division Report 2014-27 (TR).
- E. Granger, D. Gorodnichy, E. Choy, W. Khreich, P. Radtke, J. Bergeron, and D. Bissessar, "Results from evaluation of three commercial off-the-shelf face recognition systems on Chokepoint dataset", Border Technology Division, Division Report 2014-29 (TR).
- S. Matwin, D. Gorodnichy, and E. Granger, "Using smooth ROC method for evaluation and decision making in biometric systems", Border Technology Division, Division Report 2014-10 (TR).
- D. Gorodnichy, E. Granger, S. Matwin, and E. Neves, "3D face generation tool Candide for better face matching in surveillance video", Border Technology Division, Division Report 2014-11 (TR).
- E. Neves, S. Matwin, D. Gorodnichy, and E. Granger, "Evaluation of different features for face recognition in video", Border Technology Division, Division Report 2014-31 (TR).

The PROVE-IT(FRiV) project took place from August 2011 till March 2013. This document was drafted and discussed with project partners in March 2013 at the Video Technology for National Security (VT4NS) forum. The final version of it was produced in June 2014.

Typesetting: All tabulated content in this report was produced automatically using LATEX content for improved source control, flexibility and maintainability. The report contains automatically generated hyperlink references and table of contents for easier navigation and reading on-line.

Contact: Correspondence regarding this report should be directed to DMITRY dot GORODNICHY at CBSA dot GC dot CA.

Contents

List of Figures

List of Tables

1 Introduction

This report is a supplement to Division Report 2014-27 (TR) ''Evaluation methodology for face recognition technology in video surveillance applications" (by E. Granger and D. Gorodnichy) [1]. It presents the evaluation results of three widely deployed Face Recognition (FR) Commercial Off-The-Shelf (COTS) systems: Cognitec (FaceVACS-SDK 8.5 Release Date: 2011-12-19), PittPatt (Face Detection, Tracking and Recognition FTR SDK 5.2.2, Release Date: 2010) and Neurotechnology (Verilook SDK 5.4, Release Date: 2011). A multi-level performance analysis introduced in [1] is performed to evaluate and compare the system performance to one another. Four levels of details are examined according to the methodology for each target person from the Checkpoint data-set: Level 0 or score-based analysis illustrates the probability distribution of the genuine scores against that of the impostors at different face resolutions, which visually illustrates the discrimination power of the system for each target individual. Level 1 or transaction-based analysis provides the averaged description of the system in terms of false positive and false negative rates aggregated over all transactions, expressed using Receiver Operative Curves (ROC), Detection Error Trade-off (DET), and Precision-Recall Operating Characteristic (PROC) curves. Level 2 or subject-based analysis describes the performance of the system using the-so-called "Doddington's Zoo" categorization of individuals, which detects whether an individual belongs to an easier or a harder classes of people that the system is able to recognize [2, 3]. Finally, Level 3 or temporal analysis allows one to examine the overall discrimination power of the system by accumulating the positive predictions while tracking a person over time and computing the recognition confidence based thereon. Two-page Report Cards summarizing the performance of the system for each target individual are published, thus providing an exhaustive report of the systems performance on a variety of different target subjects. As highlighted in previous report [1] and other publications [4, 5], such exhaustive reporting of the biometric system performance is required when the variation of system performance from one target individual to another is suspected. Our results show that this is indeed the case for all three tested COTS FR systems.

2 Chokepoint dataset

System performance is evaluated using portal 2 data from the Chokepoint dataset [6], which simulates the Type 2 surveillance environments similar to those observed in airports [1] where individuals pass in a natural free-flow way in a narrow corridor.

To capture the data, the array of three cameras is mounted above a door, used for simultaneously recording the entry of a person from three viewpoints (see Figure 1).

The data consists of 25 subjects (19 male and 6 female) in portal 1, and 29 subjects (23 male and 6 female) in portal 2. Videos were recorded over two sessions 1 month apart. In total, it consists of 54 video sequences and 64,204 labeled face images. Each sequence was named according to the recording conditions, where P, S, and C stand for portal, sequence and camera, respectively. E and L indicate subjects either entering or leaving the portal. Frames were captured with the 3 cameras at 30 fps with an SVGA resolution (800X600 pixels), and faces incorporate variations of illumination, expression, pose, occlusion,

sharpness and misalignment due to automatic frontal detection.

In the test sequences, 29 known individual walk through a chokepoint for a total of 1281 events.

As discussed in [1], the Chokepoint data set is suitable for medium- to large-scale benchmarking of systems for mono-modal recognition and tracking of faces over one or more cameras in watch-list applications. It is provided with the ground truth (person ID, eye location and ROIs for each frame), as well as a high resolution mug shot for each individual in the data set (see Figure 3)

Table 1 presents a list of the Chokepoint video sequences used for evaluations. Figure 2 shows some frames for target individual 1 from the Chokepoint sequence P2L-S4-C1.1.

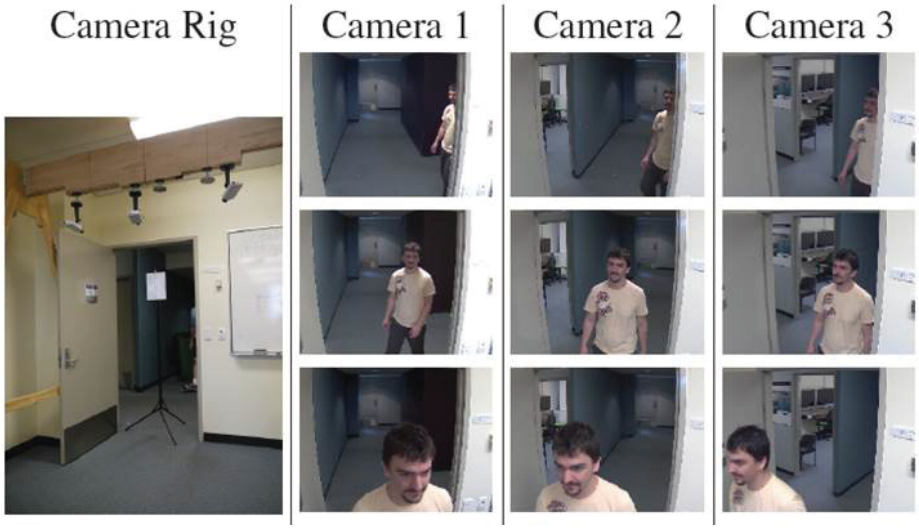

Figure 1: Setup used to capture the Chokepoint video data set.

For the testing, ten individuals are randomly selected from the Chokepoint dataset as target individuals and included in the watch list. The target individuals include six males and four females, and have the following identification numbers (id) in the Chokepoint dataset: $1, 4, 5, 7, 9, 10, 11, 12, 16$ and 29. For each

Figure 2: Frames corresponding to individual #1 from the Chokepoint data set sequence P2L-S4-C1.1.

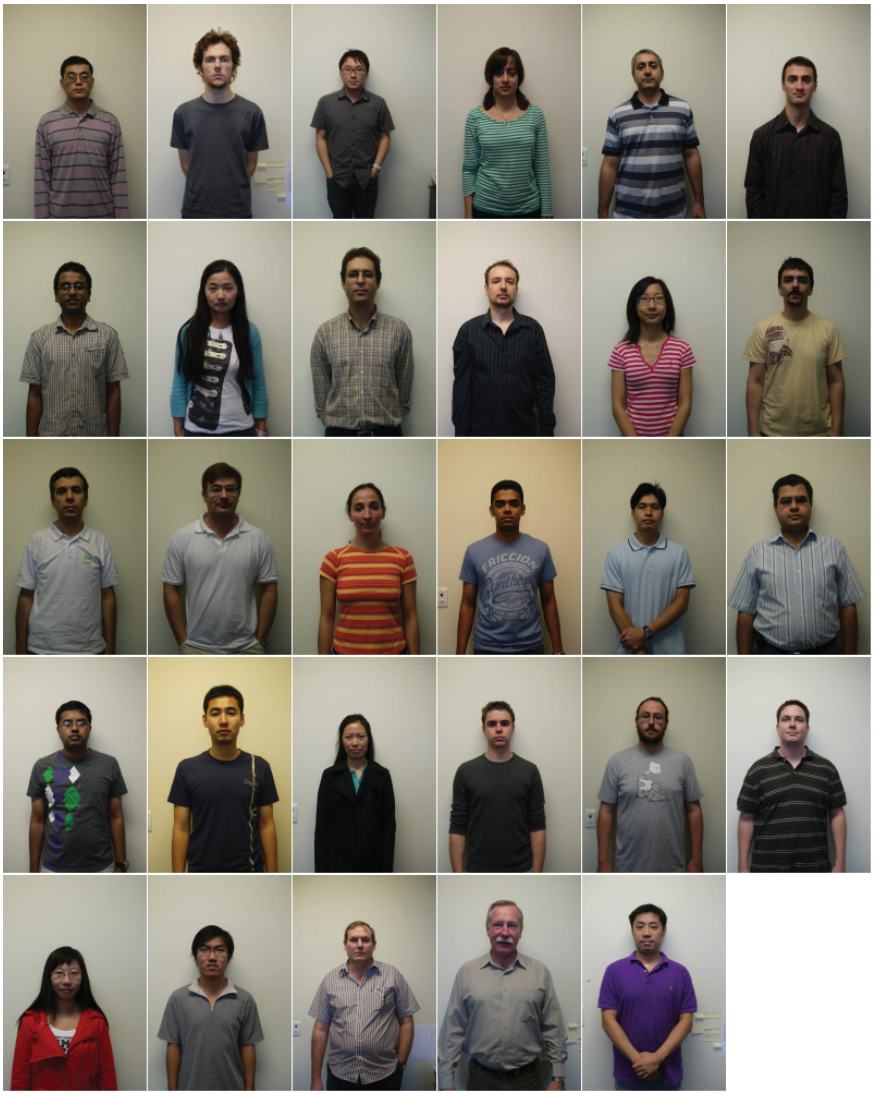

Figure 3: Gallery of still images used to create a watch list in the Chokepoint data set.

target individual, the remaining 28 individuals are considered as impostors. The performance is evaluated based on a fixed operating point (face matching threshold) of 5% false positive rate, using three different distances between the eyes: $10, 20$ and 30 pixels. The video streams from Chokepoint dataset portal 2, session 1, camera 1.1 (P2L_S1_C1.1) are considered as a validation set and used to compute matching thresholds for each target individual and each distance between the eyes. These thresholds are then applied to Chokepoint dataset portal 2, session 4, camera 1.1 (P2L_S4_C1.1), to evaluate systems performance for

Table 1: Chokepoint video sequences selected for performance evaluation. The sequences are captured with one of three cameras when subject are leaving portal 2. In all sequences, only one person passes the portal at a time.

Data sequences	no. of subjects	type of scenario
1) P2L_S1_C1, P2L_S1_C2, P2L_S1_C3	1	type 1, with different cameras
2) P2L_S2_C1, P2L_S3_C1, P2L_S4_C1	1	type 1, with different recorded sequence
3) P2L_S4_C1, P2L_S4_C2, P2L_S5_C3	24, crowded	type 2, with different cameras

each target individual.

As emphasized in [1] and further illustrated in this report, user-specific or template-specific thresholds allow for improved system performance since some individuals are naturally more difficult to recognize than others, and the risk associated with recognition errors varies from one individual to another. However, setting (or optimizing) a specific threshold for each target individual makes systems comparison more difficult to illustrate and harder to summarize. In particular, averaging the performances among all target individuals (each with a different matching threshold) may not provide meaningful results. Therefore, the detailed results for each system that are based on each target individual as well as on the distance between the eyes are presented as report cards in special Appendix of this report. For each system, the report cards illustrate the four levels of performance analysis for each individual in the watch list according to the three distances between the eyes.

3 Conclusions

This report presented the results obtained from the evaluation of three COTS FR products (Cognitec, PittPatt and NeuroTechnology) on the publicly available Chokepoint data-set following the evaluation methodology and protocol defined in [1]. The results are presented using the two-page Reports Cards, which are generated for each of ten randomly chosen target individuals from the Chokepoint data-set. These results summarize the ability of the system to automatically detect and recognize target individuals in a surveillance video-stream, while highlighting their strengths and vulnerabilities.

The obtained evaluation results reported here and in report [1], along with the survey of academic and commercial state of art solutions presented in separate reports [7, 8], provided the basis for the assessment of the readiness of the FR technology for video surveillance applications, which was the key objective of the PROVE-IT(FRiV) study and which was reported in [9, 10] and further refined in [11].

References

[1] E. Granger and D. Gorodnichy, "Evaluation methodology for face recognition technology in video surveillance applications," CBSA, Tech. Rep. 2014-27 (TR), 2014.

[2] G. Doddington, W. Liggett, A. Martin, M. Przybocki, and D. Reynolds, "Sheep, goats, lambs and wolves: A statistical analysis of speaker performance in the nist 1998 speaker recognition evaluation," in *International Conference on Spoken Language Processing*, 1998.

[3] A. Rattani, G. Marcialis, and F. Roli, "An experimental analysis of the relationship between biometric template update and the doddingtons zoo: A case study in face verification," in *Image Analysis and Processing ICIAP 2009*, ser. Lecture Notes in Computer Science, P. Foggia, C. Sansone, and M. Vento, Eds. Springer Berlin / Heidelberg, 2009, vol. 5716, pp. 434–442.

[4] N. Yager and T. Dunstone, "The biometric menagerie," *IEEE Transactions on Pattern Recognition and Machine Intelligence*, pp. 220–230, 2010.

[5] B. DeCann and A. Ross, "Relating roc and cmc curves via the biometric menagerie," in *Proc. of 6th IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS), (Washington DC, USA), September 2013*.

[6] Y. Wong, S. Chen, S. Mau, C. Sanderson, and B. C. Lovell, "Patch-based probabilistic image quality assessment for face selection and improved video-based face recognition," *Computer Vision and Pattern Recognition Workshops*. [Online]. Available: http://itee.uq.edu.au/~uqywong6/chokepoint.html

[7] D. Gorodnichy, E. Granger, and P.Radtke, "Survey of commercial technologies for face recognition in video," CBSA, Border Technology Division, Tech. Rep. 2014-22 (TR), 2014.

[8] E. Granger, P. Radtke, and D. Gorodnichy, "Survey of academic research and prototypes for face recognition in video," CBSA, Border Technology Division, Tech. Rep. 2014-25 (TR), 2014.

[9] D. Gorodnichy and E. Granger, "Evaluation of face recognition for video surveillance," *Proceedings of NIST International Biometrics Performance Conference (IBPC 2012)*.

[10] D. Bissessar, E. Choy, D. Gorodnichy, and T. Mungham, "Face recognition and event detection in video: An overview of prove-it projects (biom401 and bts402)," *CBSA Science and Engineering Directorate Technical Report 2013-04, June 2013*.

[11] D. Gorodnichy, E. Granger, J.-P. B. abd D. Bissessar, E. Choy, T. Mungham, R. Laganiere, S. Matwin, E. Neves, C. Pagano, M. D. la Torre, and P. Radtke, "Prove-it(friv): framework and results," *Proceedings of NIST International Biometrics Performance Conference (IBPC 2014)*.

4 Evaluation Results for Cognitec System

Biometric System: Cognitec	**Biometric Modality**: Face
Data Source: Chokepoint – Portal 2, leaving portal setup, session 4, camera 1.1	**Individual Template**: 1

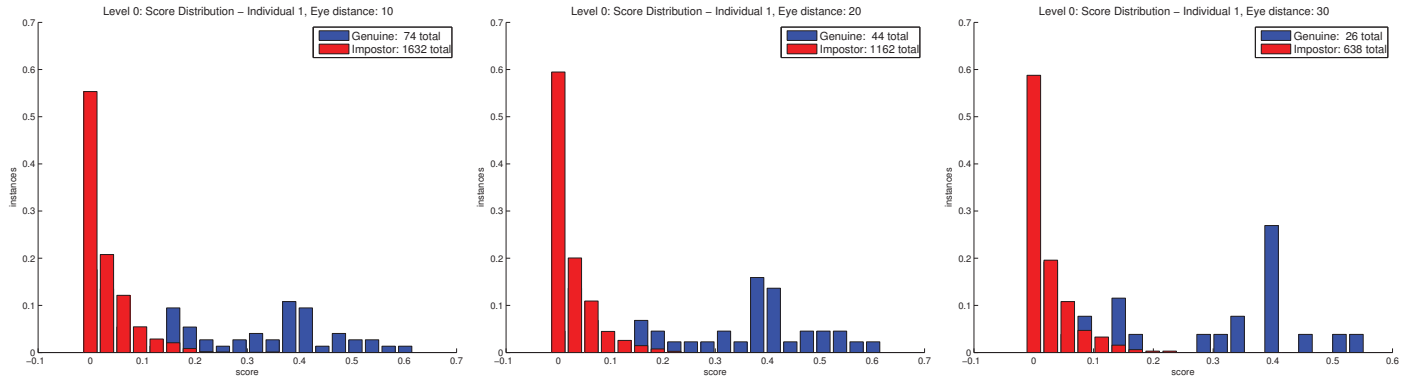

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

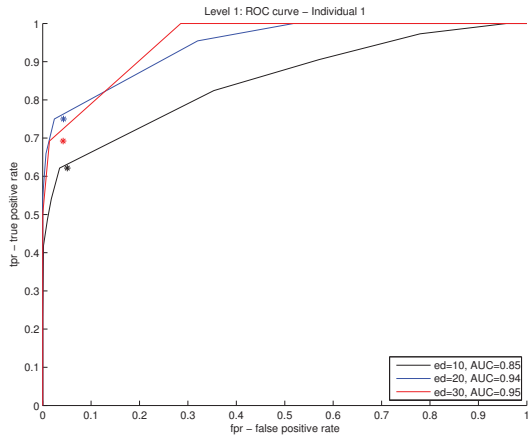

Fig. 2: ROC curve.

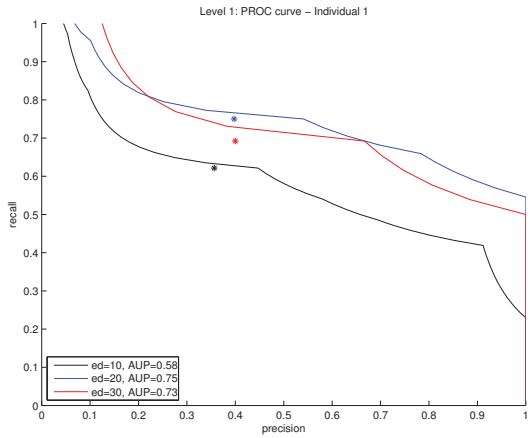

Fig. 3: PROC curve.

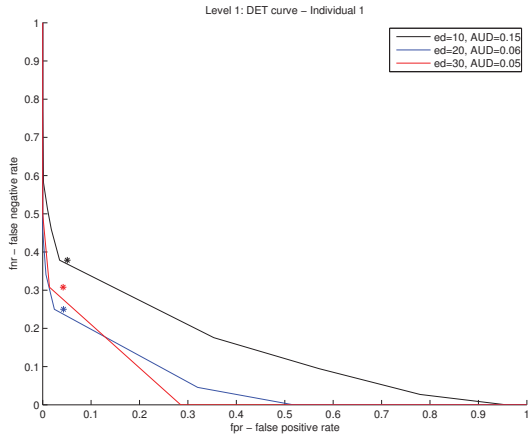

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	74	44	26
Impostor faces (total)	1632	1162	638
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1383	0.1315	0.1294
False positive rates	5.09%	4.30%	4.23%
True positive rates	62.16%	75.00%	69.23%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	**62.16**%	0.00%	0.00%	1.61%	8.20%	1.18%	14.47%	0.00%	3.39%	3.03%	1.35%	3.03%	7.78%	6.33%	0.00%
20 px.	**75.00**%	0.00%	0.00%	2.63%	4.44%	1.69%	19.67%	0.00%	0.00%	4.00%	0.00%	0.00%	3.08%	5.66%	0.00%
30 px.	**69.23**%	0.00%	0.00%	0.00%	8.33%	0.00%	12.50%	0.00%	0.00%	3.57%	0.00%	0.00%	5.13%	6.67%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	5.00%	3.08%	4.23%	2.78%	7.55%	0.00%	0.00%	14.75%	6.17%	3.90%	11.11%	1.37%	3.23%	3.45%	5.63%
20 px.	2.70%	0.00%	5.66%	2.00%	9.52%	0.00%	0.00%	12.50%	5.26%	1.79%	2.78%	2.04%	4.00%	4.65%	6.52%
30 px.	5.26%	0.00%	6.25%	3.70%	12.50%	0.00%	0.00%	14.29%	3.23%	3.12%	0.00%	3.85%	3.57%	7.69%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 1. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

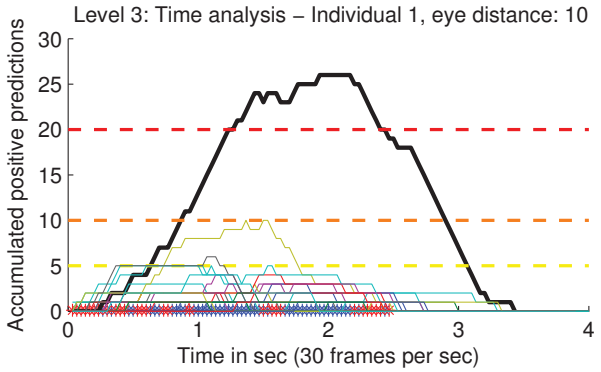

Fig. 5: Accumulated detections for 10 pixels between eyes.

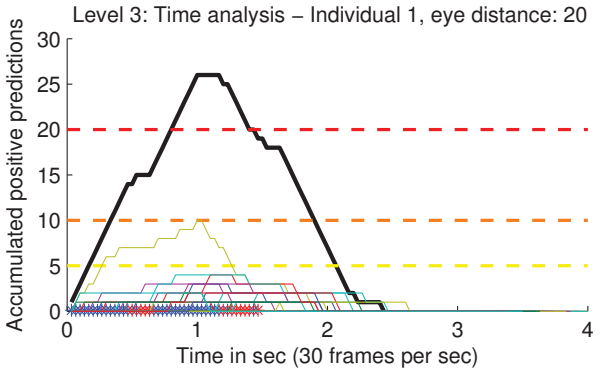

Fig. 6: Accumulated detections for 20 pixels between eyes.

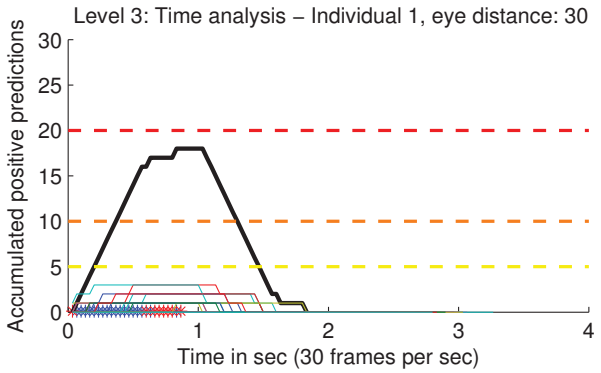

Fig. 7: Accumulated detections for 30 pixels between eyes.

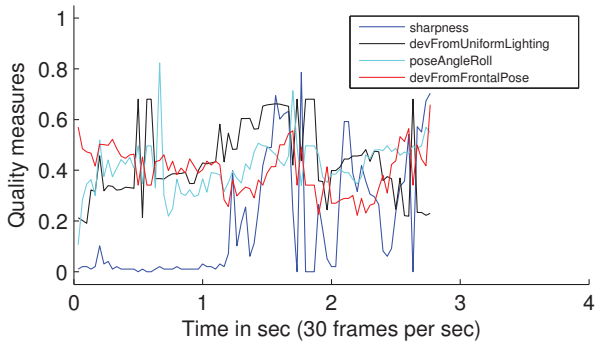

Fig. 8: Variations of quality measures.

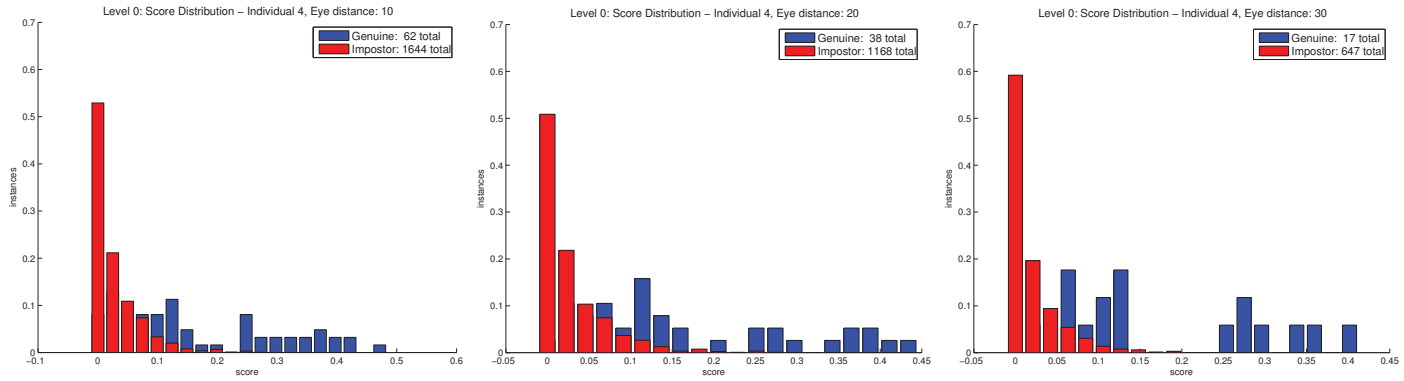

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

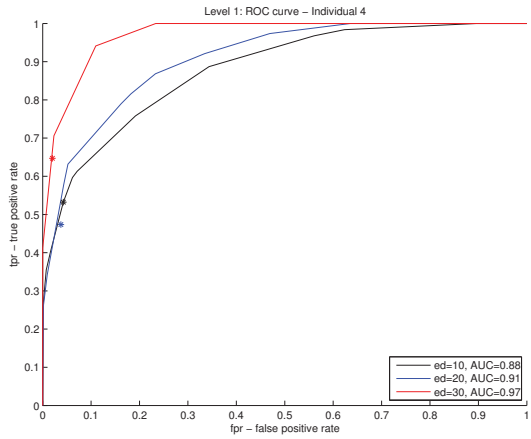

Fig. 2: ROC curve.

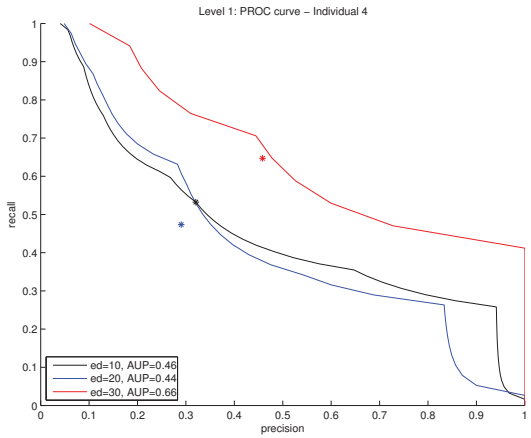

Fig. 3: PROC curve.

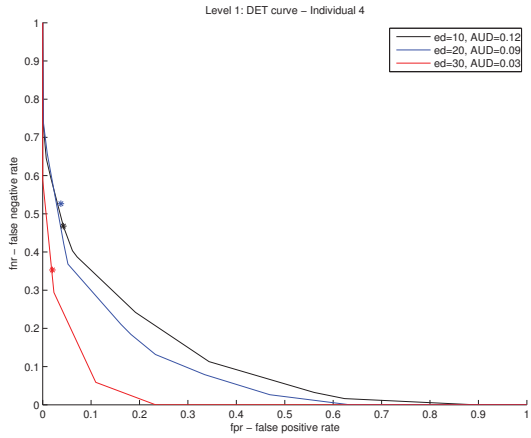

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	62	38	17
Impostor faces (total)	1644	1168	647
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1248	0.1305	0.1225
False positive rates	4.26%	3.77%	2.01%
True positive rates	53.23%	47.37%	64.71%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	1.79%	0.00%	**53.23%**	6.56%	3.53%	10.53%	0.00%	3.39%	1.52%	6.76%	3.03%	0.00%	5.06%	0.00%
20 px.	0.00%	2.17%	0.00%	**47.37%**	6.67%	1.69%	6.56%	0.00%	2.63%	0.00%	9.62%	0.00%	0.00%	1.89%	0.00%
30 px.	0.00%	0.00%	0.00%	**64.71%**	8.33%	3.03%	3.12%	0.00%	0.00%	0.00%	3.33%	0.00%	0.00%	3.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	4.62%	0.00%	5.56%	9.43%	0.00%	0.00%	6.56%	3.70%	5.19%	3.70%	2.74%	4.84%	6.90%	8.45%
20 px.	0.00%	3.92%	0.00%	2.00%	7.14%	0.00%	0.00%	10.00%	1.75%	7.14%	5.56%	4.08%	4.00%	9.30%	6.52%
30 px.	0.00%	0.00%	0.00%	0.00%	16.67%	0.00%	0.00%	4.76%	0.00%	0.00%	0.00%	0.00%	0.00%	3.85%	4.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 4. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

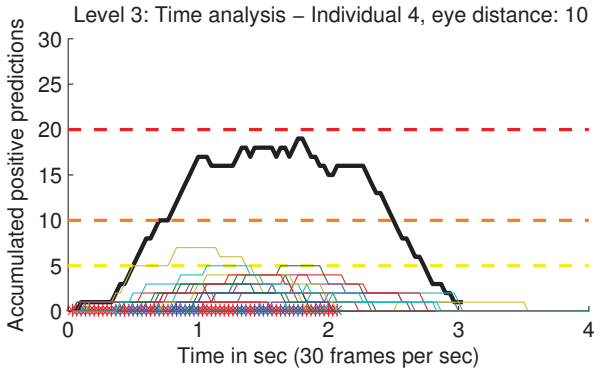

Fig. 5: Accumulated detections for 10 pixels between eyes.

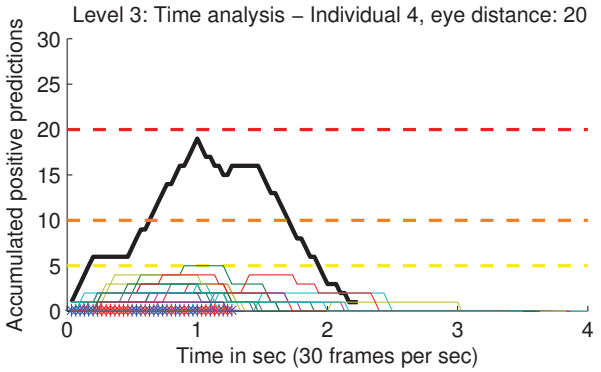

Fig. 6: Accumulated detections for 20 pixels between eyes.

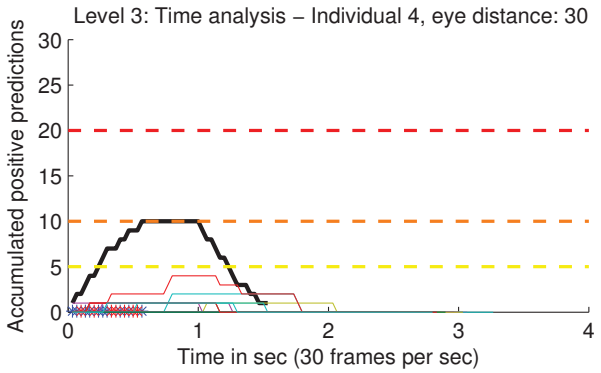

Fig. 7: Accumulated detections for 30 pixels between eyes.

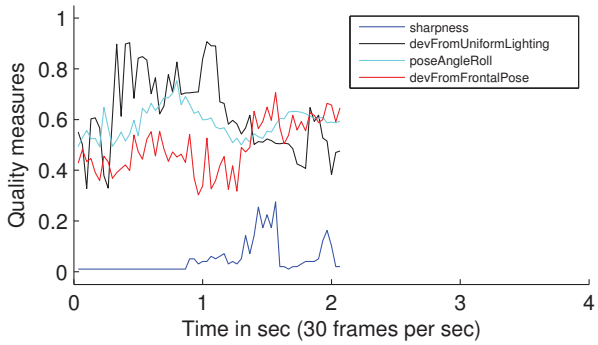

Fig. 8: Variations of quality measures.

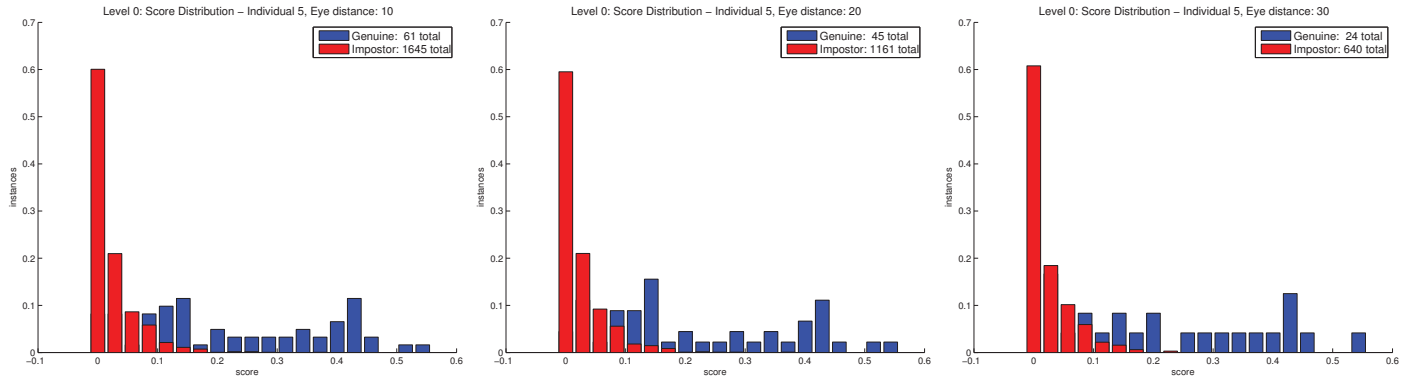

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

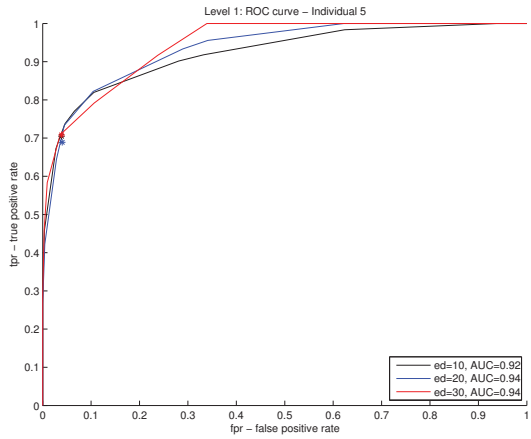

Fig. 2: ROC curve.

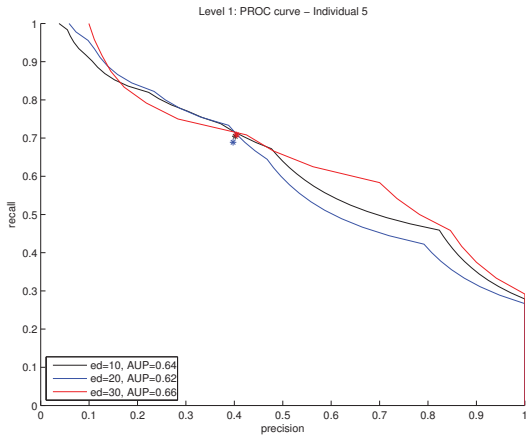

Fig. 3: PROC curve.

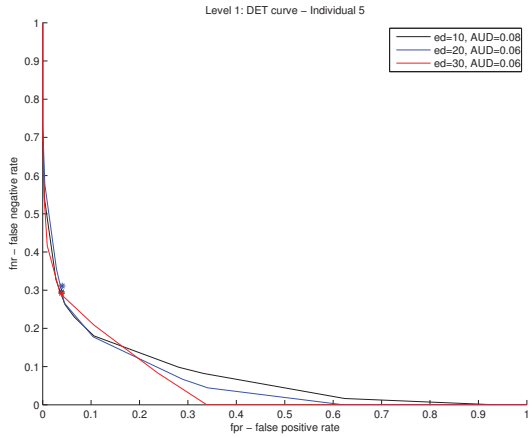

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	61	45	24
Impostor faces (total)	1645	1161	640
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1211	0.1221	0.1217
False positive rates	3.89%	4.05%	3.91%
True positive rates	70.49%	68.89%	70.83%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	6.76%	1.79%	0.00%	0.00%	**70.49%**	1.18%	5.26%	0.00%	1.69%	10.61%	0.00%	0.00%	0.00%	12.66%	0.00%
20 px.	9.09%	2.17%	0.00%	0.00%	**68.89%**	1.69%	4.92%	0.00%	0.00%	14.00%	0.00%	0.00%	0.00%	18.87%	0.00%
30 px.	15.38%	3.70%	0.00%	0.00%	**70.83%**	3.03%	9.38%	0.00%	0.00%	3.57%	0.00%	0.00%	0.00%	13.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	3.33%	0.00%	4.23%	5.56%	7.55%	0.00%	0.00%	0.00%	3.70%	2.60%	3.70%	6.85%	3.23%	6.90%	5.63%
20 px.	2.70%	0.00%	3.77%	4.00%	7.14%	0.00%	0.00%	0.00%	3.51%	0.00%	0.00%	8.16%	4.00%	4.65%	6.52%
30 px.	0.00%	0.00%	3.12%	7.41%	12.50%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%	0.00%	7.69%	4.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 5. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

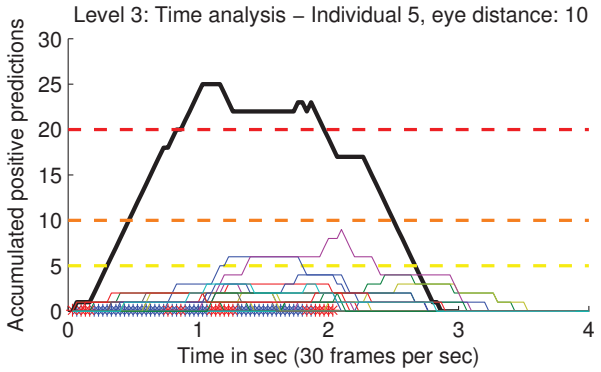

Fig. 5: Accumulated detections for 10 pixels between eyes.

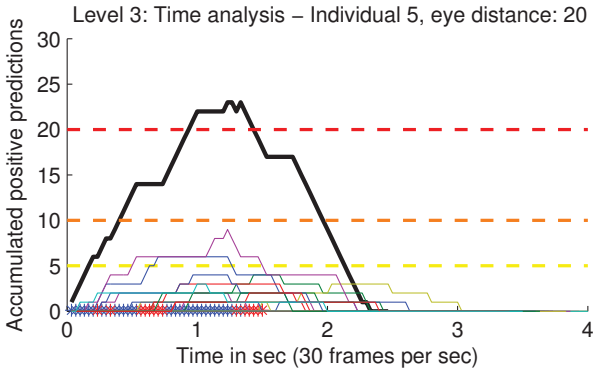

Fig. 6: Accumulated detections for 20 pixels between eyes.

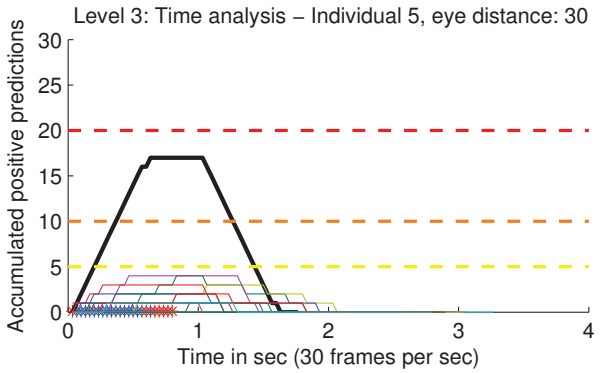

Fig. 7: Accumulated detections for 30 pixels between eyes.

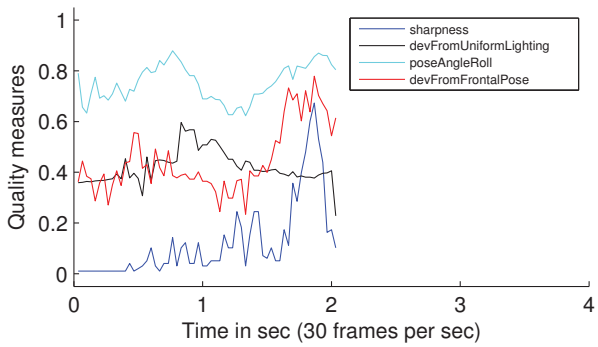

Fig. 8: Variations of quality measures.

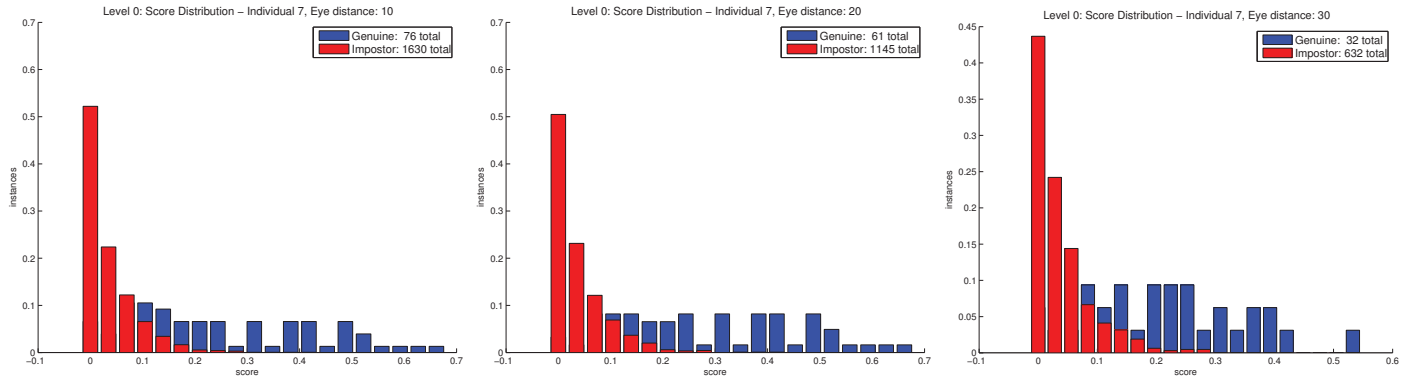

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

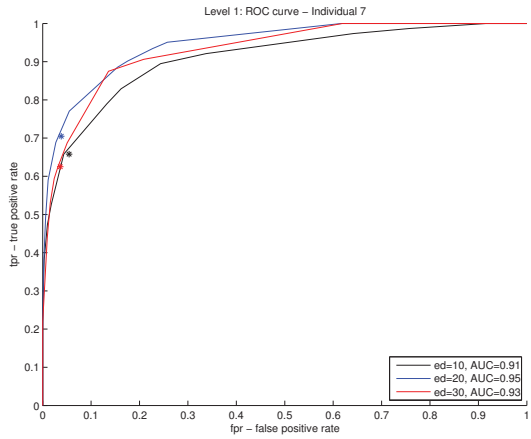

Fig. 2: ROC curve.

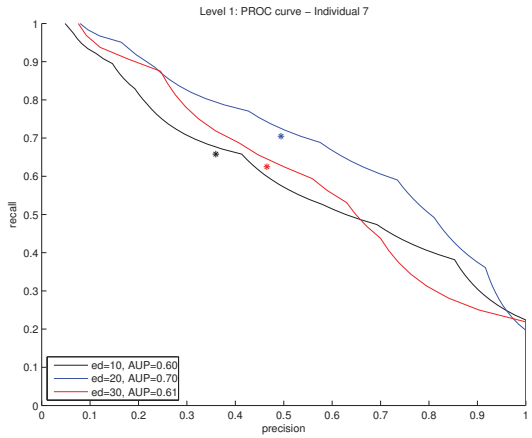

Fig. 3: PROC curve.

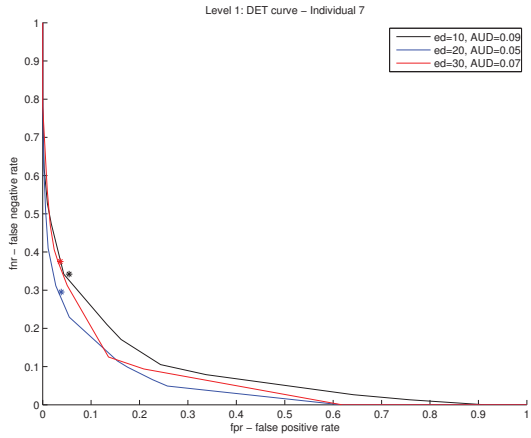

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	76	61	32
Impostor faces (total)	1630	1145	632
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1501	0.1728	0.1695
False positive rates	5.46%	3.84%	3.64%
True positive rates	65.79%	70.49%	62.50%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	2.70%	3.57%	0.00%	1.61%	9.84%	1.18%	**65.79%**	0.00%	0.00%	13.64%	5.41%	9.09%	6.67%	7.59%	0.00%
20 px.	0.00%	2.17%	0.00%	0.00%	4.44%	0.00%	**70.49%**	0.00%	0.00%	18.00%	3.85%	8.89%	3.08%	7.55%	0.00%
30 px.	0.00%	3.70%	0.00%	0.00%	0.00%	0.00%	**62.50%**	0.00%	0.00%	7.14%	6.67%	13.04%	5.13%	10.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	3.33%	7.69%	4.23%	9.72%	11.32%	0.00%	0.00%	0.00%	3.70%	7.79%	1.85%	5.48%	6.45%	5.17%	2.82%
20 px.	0.00%	5.88%	3.77%	8.00%	7.14%	0.00%	0.00%	0.00%	1.75%	3.57%	0.00%	6.12%	0.00%	2.33%	2.17%
30 px.	0.00%	3.57%	3.12%	7.41%	4.17%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%	11.54%	0.00%	0.00%	4.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 7. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

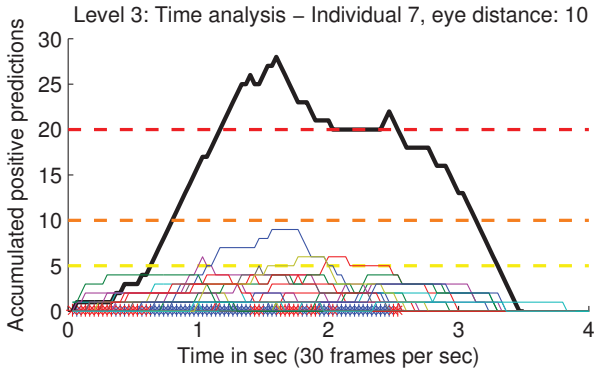

Fig. 5: Accumulated detections for 10 pixels between eyes.

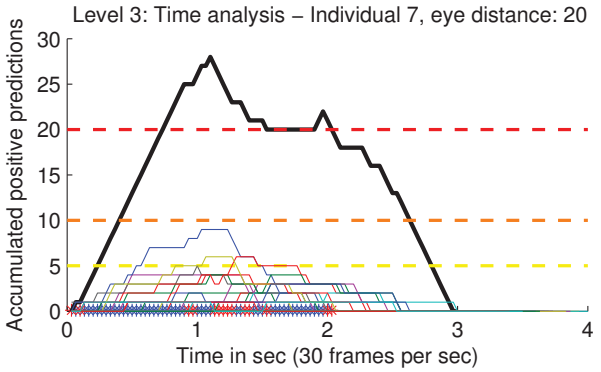

Fig. 6: Accumulated detections for 20 pixels between eyes.

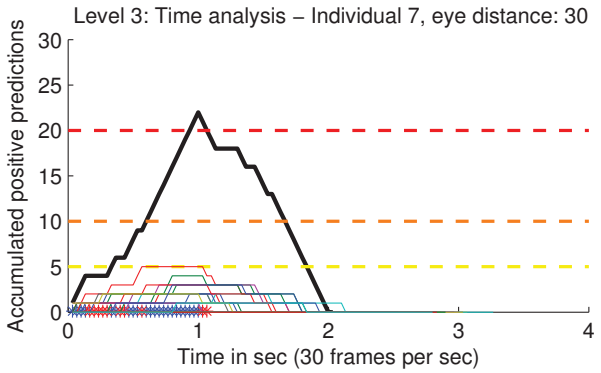

Fig. 7: Accumulated detections for 30 pixels between eyes.

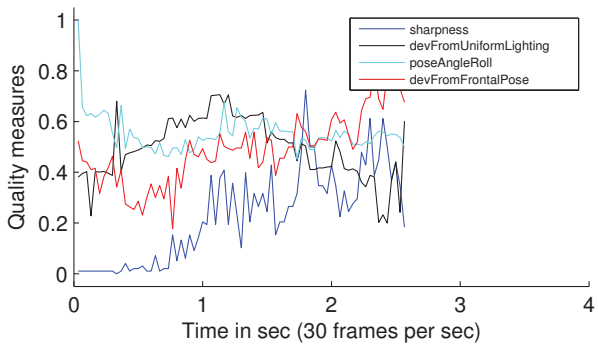

Fig. 8: Variations of quality measures.

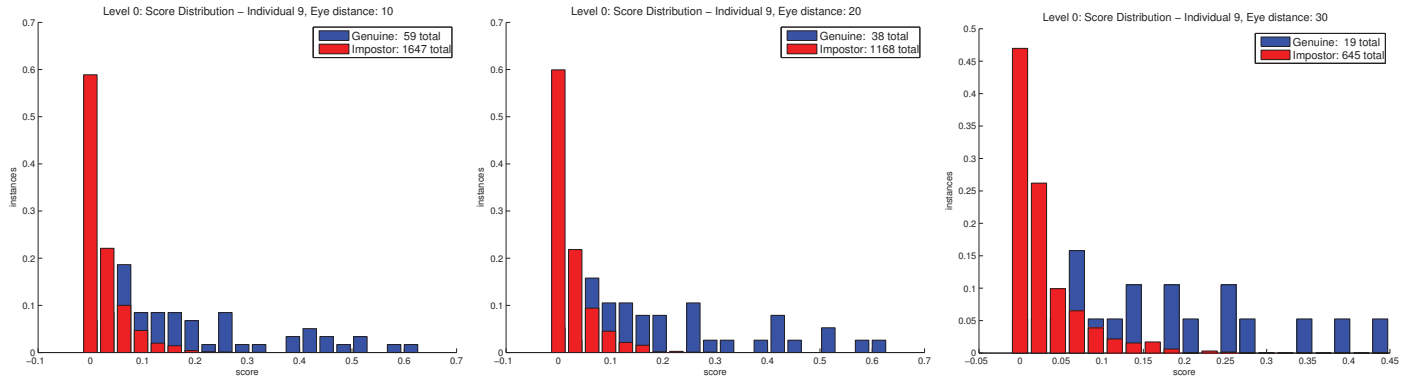

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

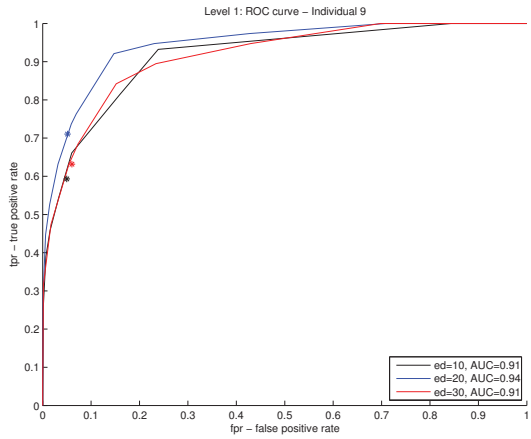

Fig. 2: ROC curve.

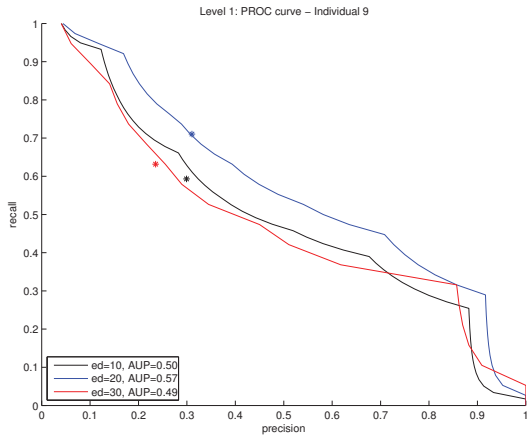

Fig. 3: PROC curve.

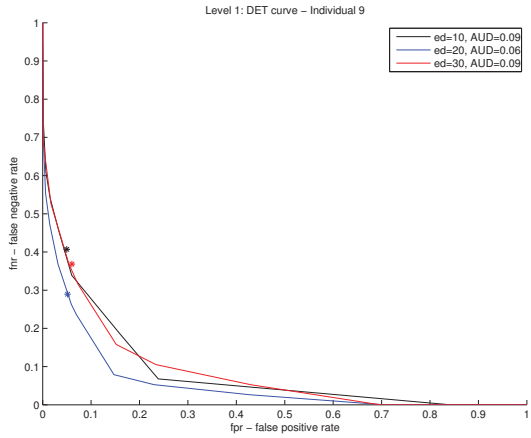

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	59	38	19
Impostor faces (total)	1647	1168	645
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1245	0.1227	0.1233
False positive rates	4.98%	5.14%	6.05%
True positive rates	59.32%	71.05%	63.16%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	1.35%	1.79%	0.00%	9.68%	9.84%	5.88%	6.58%	0.00%	**59.32%**	0.00%	10.81%	3.03%	7.78%	0.00%	0.00%
20 px.	0.00%	2.17%	0.00%	7.89%	6.67%	8.47%	8.20%	0.00%	**71.05%**	0.00%	9.62%	4.44%	7.69%	0.00%	0.00%
30 px.	0.00%	3.70%	0.00%	5.88%	0.00%	15.15%	3.12%	0.00%	**63.16%**	0.00%	16.67%	4.35%	12.82%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	8.33%	6.15%	7.04%	5.56%	1.89%	0.00%	0.00%	4.92%	2.47%	1.30%	11.11%	1.37%	0.00%	1.72%	11.27%
20 px.	5.41%	7.84%	7.55%	8.00%	2.38%	0.00%	0.00%	7.50%	1.75%	1.79%	13.89%	2.04%	0.00%	0.00%	10.87%
30 px.	5.26%	14.29%	9.38%	14.81%	0.00%	0.00%	0.00%	14.29%	3.23%	0.00%	11.76%	0.00%	0.00%	0.00%	8.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 9. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

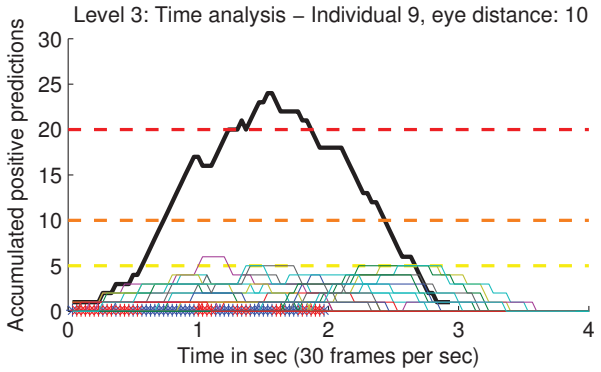

Fig. 5: Accumulated detections for 10 pixels between eyes.

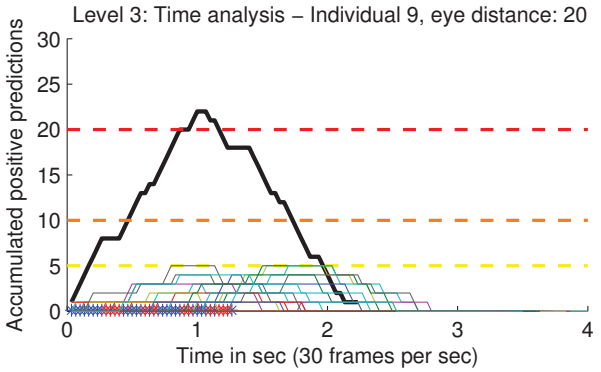

Fig. 6: Accumulated detections for 20 pixels between eyes.

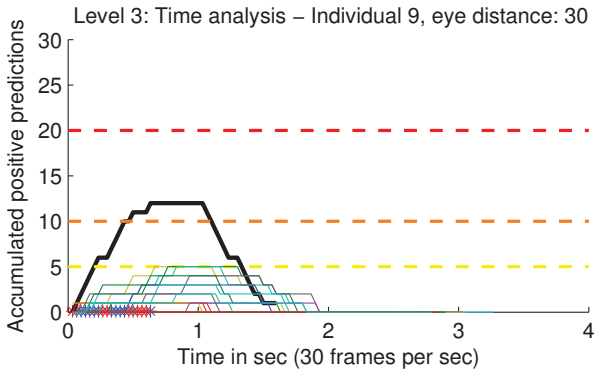

Fig. 7: Accumulated detections for 30 pixels between eyes.

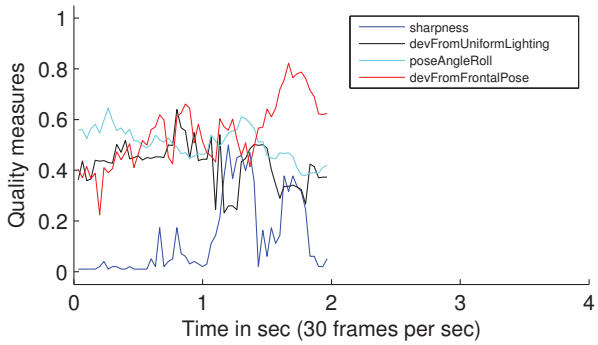

Fig. 8: Variations of quality measures.

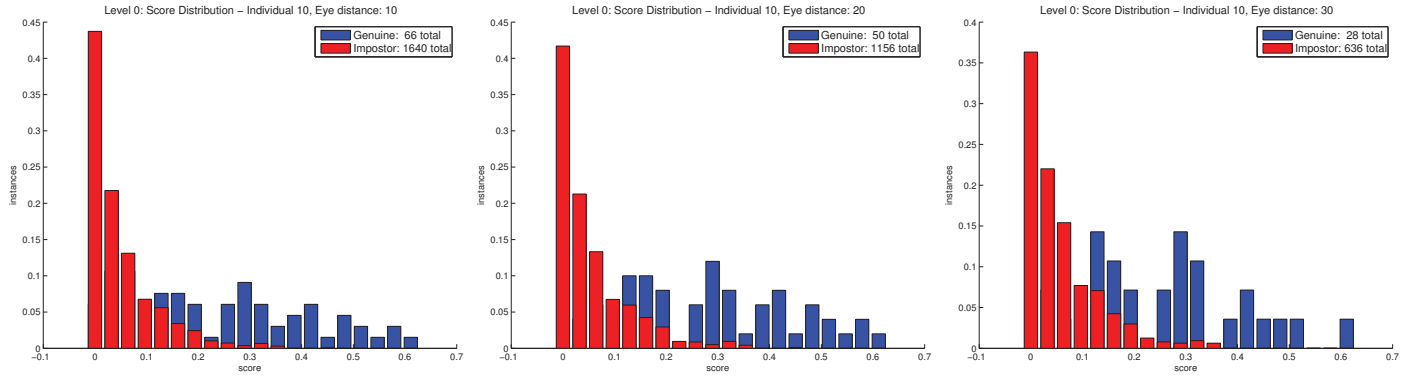

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

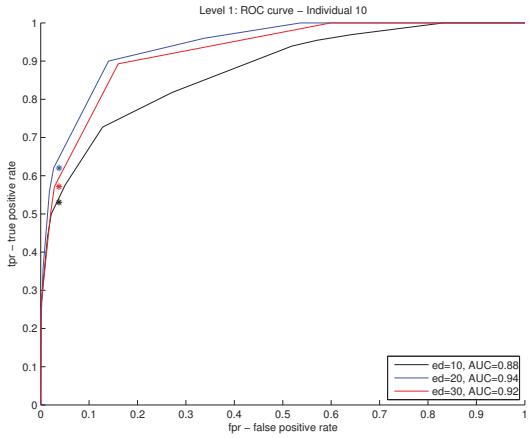

Fig. 2: ROC curve.

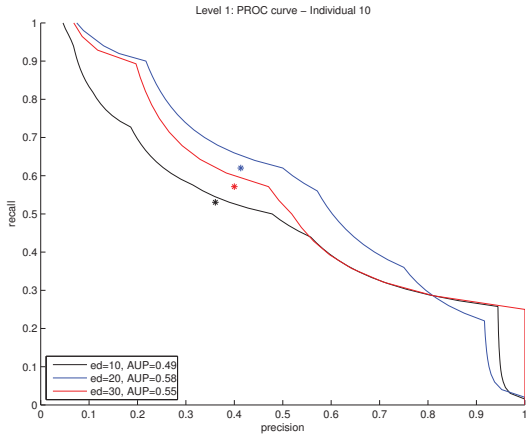

Fig. 3: PROC curve.

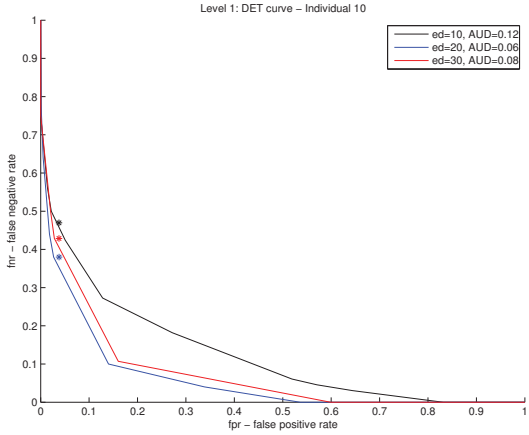

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	66	50	28
Impostor faces (total)	1640	1156	636
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.2176	0.2248	0.2507
False positive rates	3.78%	3.81%	3.77%
True positive rates	53.03%	62.00%	57.14%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	1.79%	0.00%	0.00%	6.56%	5.88%	5.26%	0.00%	0.00%	**53.03%**	0.00%	1.52%	1.11%	11.39%	0.00%
20 px.	0.00%	2.17%	0.00%	0.00%	4.44%	6.78%	3.28%	0.00%	0.00%	**62.00%**	0.00%	2.22%	1.54%	13.21%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	8.33%	9.09%	0.00%	0.00%	0.00%	**57.14%**	0.00%	4.35%	0.00%	13.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	6.15%	4.23%	1.39%	20.75%	0.00%	0.00%	6.56%	7.41%	5.19%	1.85%	1.37%	0.00%	3.45%	0.00%
20 px.	0.00%	5.88%	3.77%	0.00%	23.81%	0.00%	0.00%	5.00%	3.51%	7.14%	0.00%	2.04%	0.00%	4.65%	0.00%
30 px.	0.00%	3.57%	6.25%	0.00%	29.17%	0.00%	0.00%	0.00%	0.00%	6.25%	0.00%	0.00%	0.00%	7.69%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 10. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

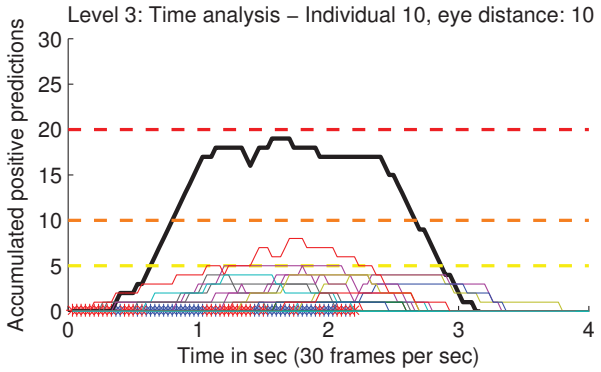

Fig. 5: Accumulated detections for 10 pixels between eyes.

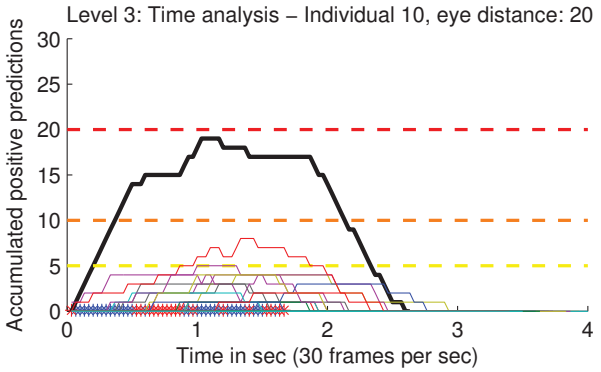

Fig. 6: Accumulated detections for 20 pixels between eyes.

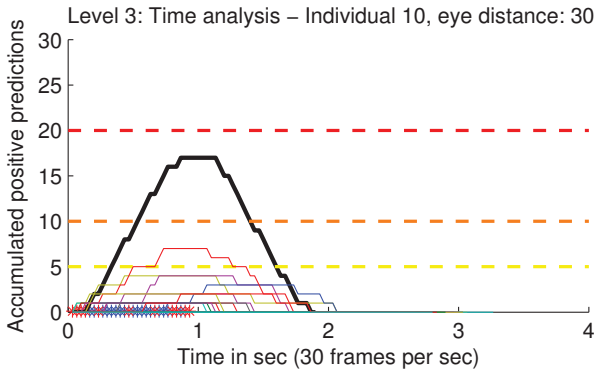

Fig. 7: Accumulated detections for 30 pixels between eyes.

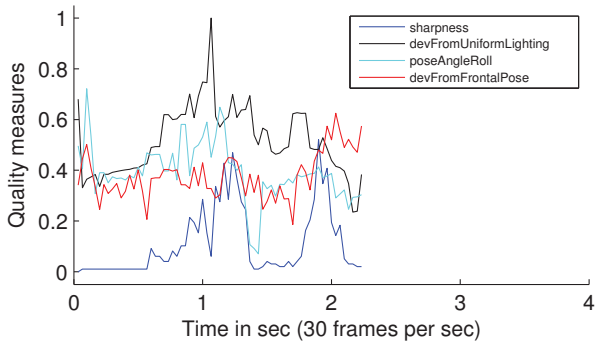

Fig. 8: Variations of quality measures.

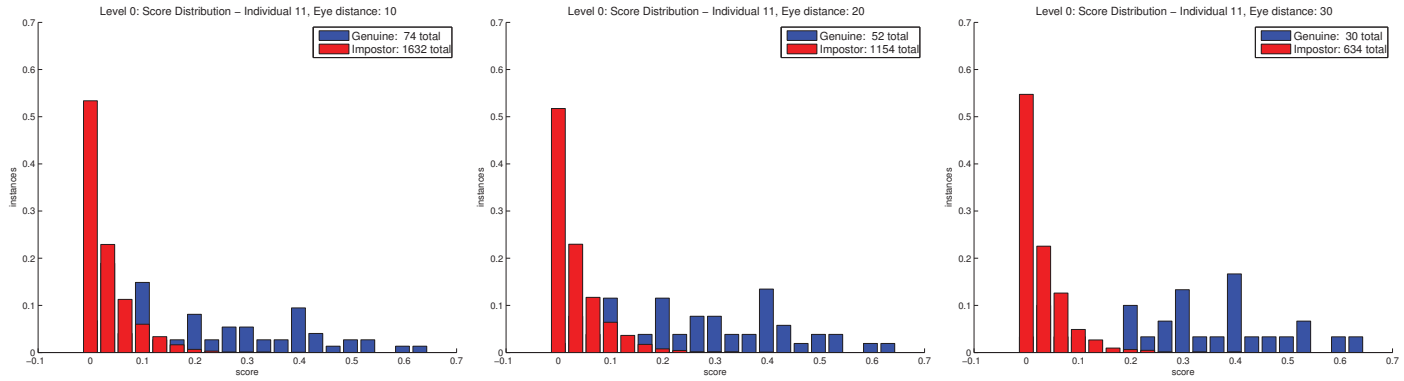

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

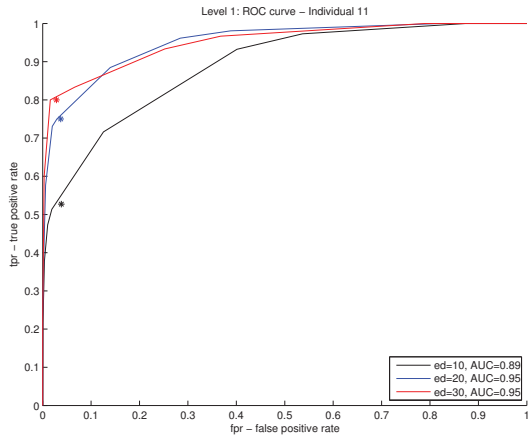

Fig. 2: ROC curve.

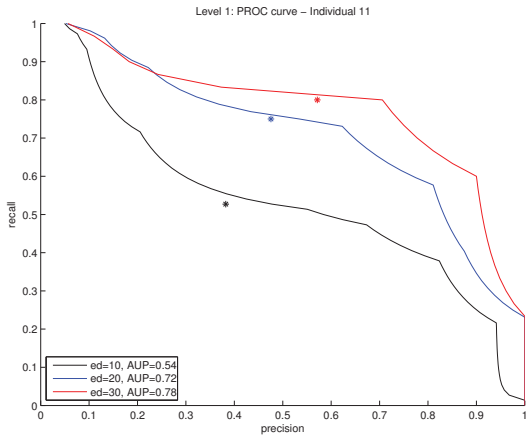

Fig. 3: PROC curve.

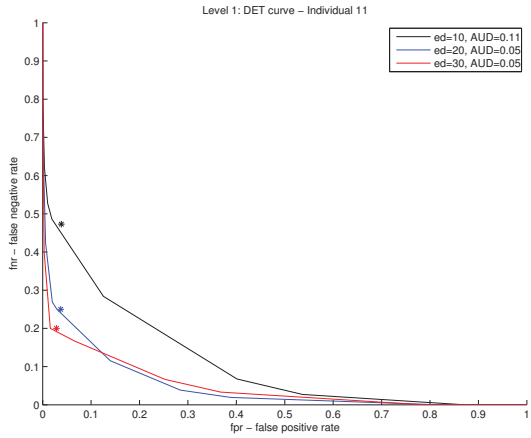

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	74	52	30
Impostor faces (total)	1632	1154	634
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1537	0.1618	0.1560
False positive rates	3.86%	3.73%	2.84%
True positive rates	52.70%	75.00%	80.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	1.35%	1.79%	0.00%	4.84%	1.64%	1.18%	3.95%	0.00%	1.69%	6.06%	**52.70%**	3.03%	7.78%	1.27%	0.00%
20 px.	0.00%	2.17%	0.00%	2.63%	0.00%	1.69%	3.28%	0.00%	0.00%	6.00%	**75.00%**	0.00%	7.69%	0.00%	0.00%
30 px.	0.00%	3.70%	0.00%	5.88%	0.00%	3.03%	0.00%	0.00%	0.00%	0.00%	**80.00%**	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	3.33%	0.00%	1.41%	1.39%	3.77%	0.00%	0.00%	4.92%	18.52%	2.60%	0.00%	0.00%	0.00%	17.24%	2.82%
20 px.	2.70%	0.00%	1.89%	2.00%	4.76%	0.00%	0.00%	7.50%	24.56%	1.79%	0.00%	0.00%	0.00%	16.28%	0.00%
30 px.	0.00%	0.00%	3.12%	0.00%	0.00%	0.00%	0.00%	0.00%	19.35%	3.12%	0.00%	0.00%	0.00%	26.92%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 11. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

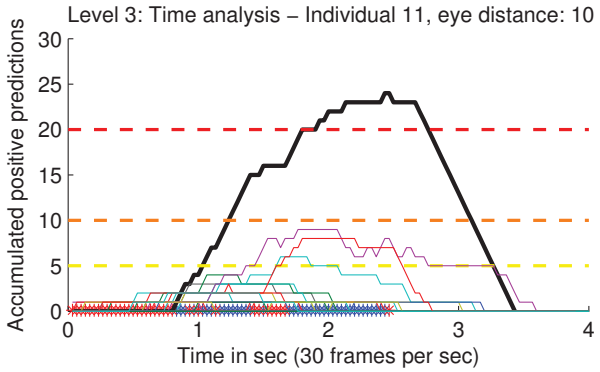

Fig. 5: Accumulated detections for 10 pixels between eyes.

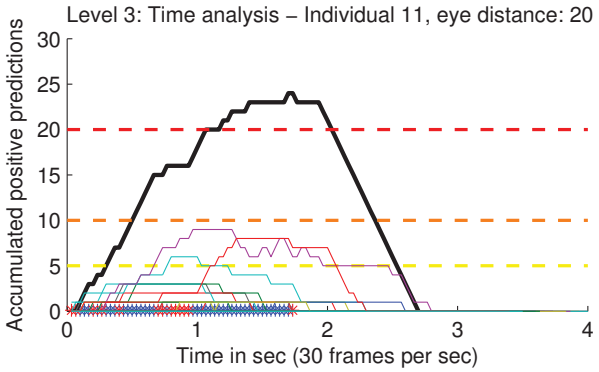

Fig. 6: Accumulated detections for 20 pixels between eyes.

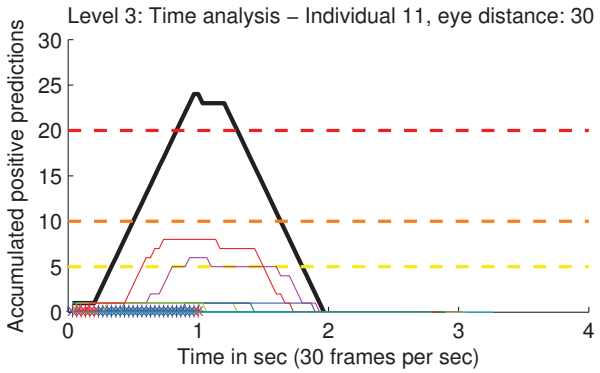

Fig. 7: Accumulated detections for 30 pixels between eyes.

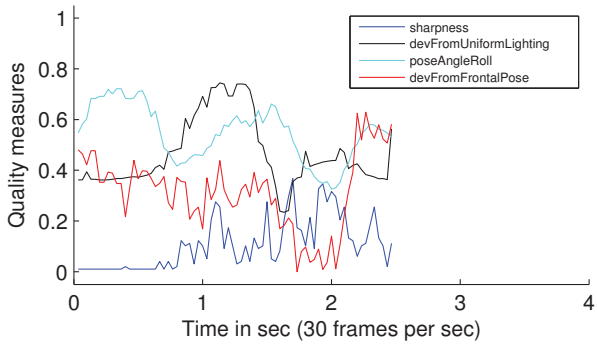

Fig. 8: Variations of quality measures.

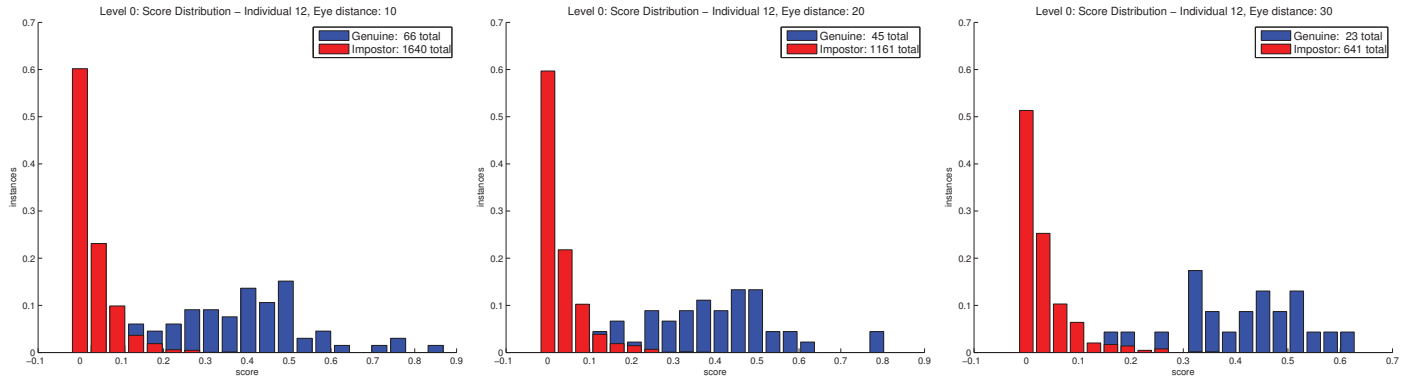

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

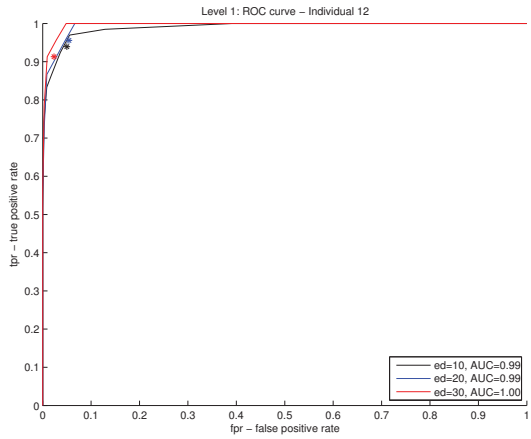

Fig. 2: ROC curve.

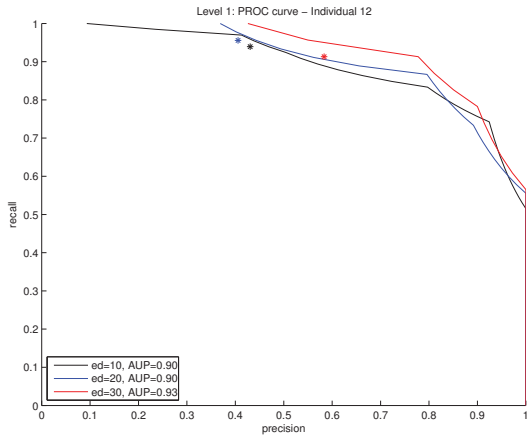

Fig. 3: PROC curve.

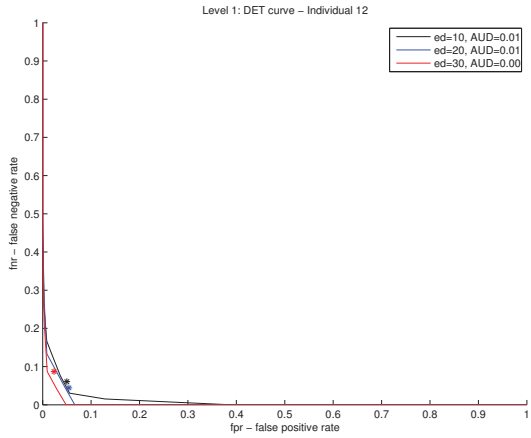

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	66	45	23
Impostor faces (total)	1640	1161	641
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1498	0.1522	0.2134
False positive rates	5.00%	5.43%	2.34%
True positive rates	93.94%	95.56%	91.30%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	9.46%	3.57%	0.00%	11.29%	22.95%	3.53%	0.00%	0.00%	1.69%	9.09%	1.35%	**93.94%**	0.00%	2.53%	0.00%
20 px.	15.91%	4.35%	0.00%	15.79%	24.44%	5.08%	0.00%	0.00%	0.00%	8.00%	1.92%	**95.56%**	0.00%	3.77%	0.00%
30 px.	19.23%	0.00%	0.00%	5.88%	20.83%	0.00%	0.00%	0.00%	0.00%	3.57%	0.00%	**91.30%**	0.00%	3.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	8.45%	6.94%	1.89%	0.00%	0.00%	3.28%	3.70%	6.49%	16.67%	8.22%	1.61%	0.00%	1.41%
20 px.	0.00%	0.00%	9.43%	4.00%	2.38%	0.00%	0.00%	5.00%	3.51%	3.57%	16.67%	12.24%	2.00%	0.00%	0.00%
30 px.	0.00%	0.00%	3.12%	0.00%	0.00%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 12. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

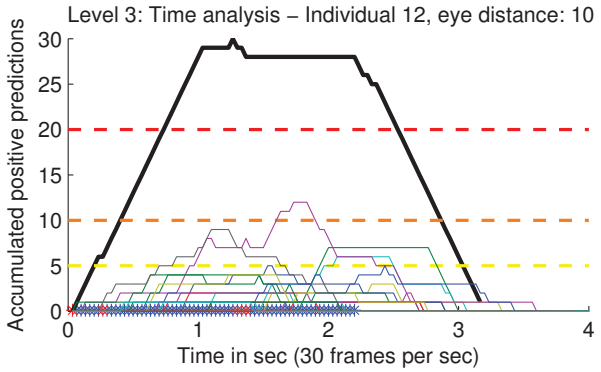

Fig. 5: Accumulated detections for 10 pixels between eyes.

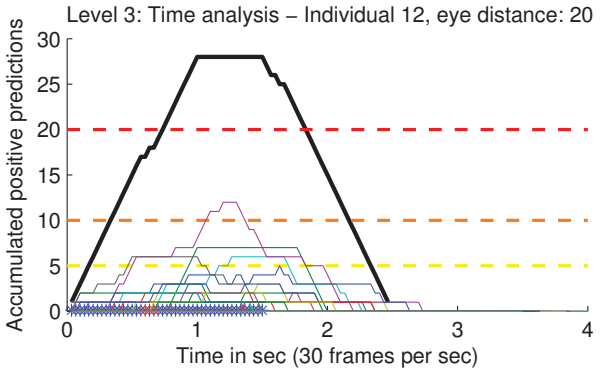

Fig. 6: Accumulated detections for 20 pixels between eyes.

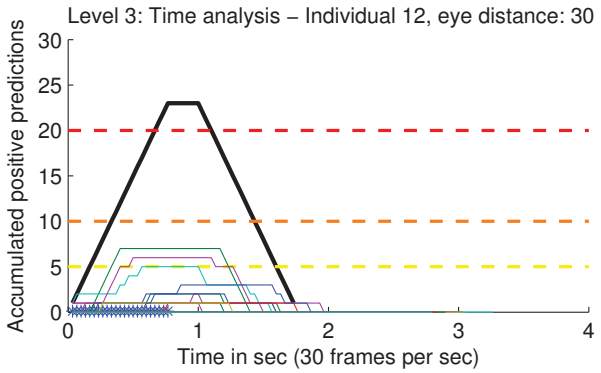

Fig. 7: Accumulated detections for 30 pixels between eyes.

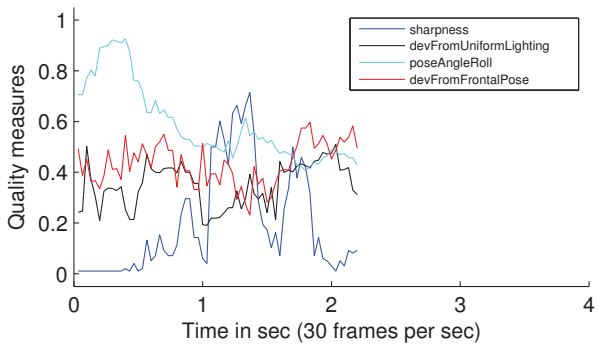

Fig. 8: Variations of quality measures.

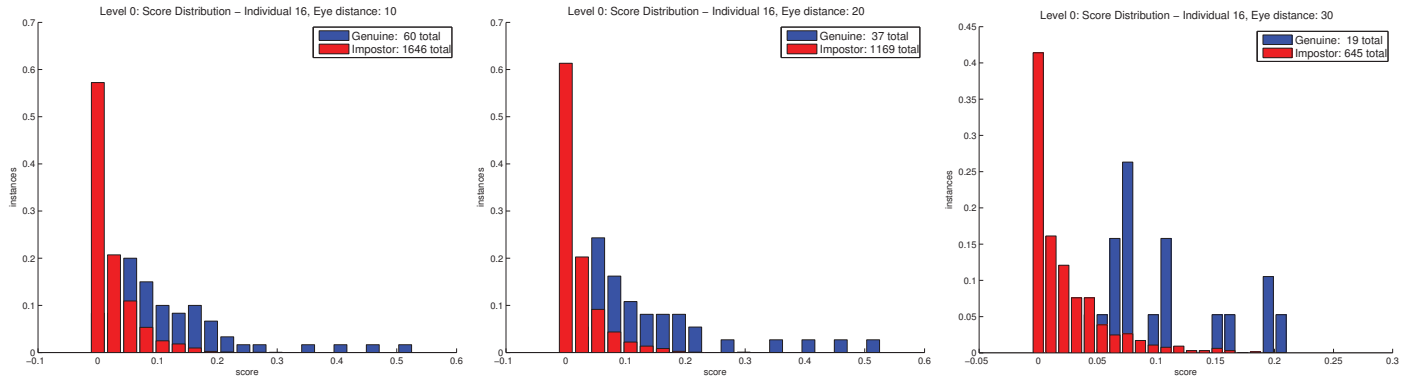

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

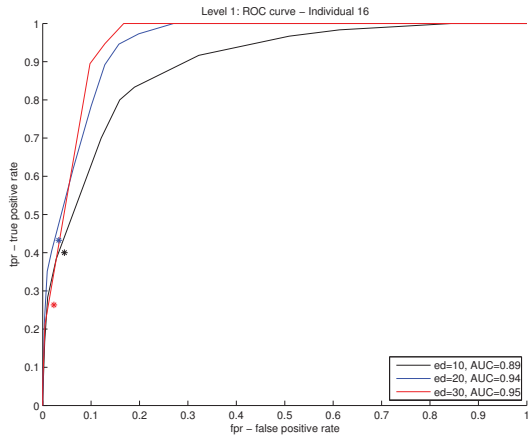

Fig. 2: ROC curve.

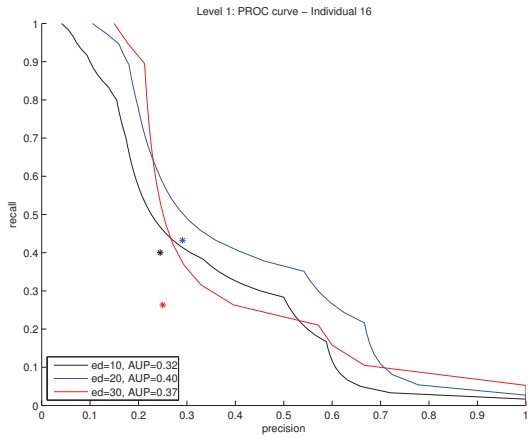

Fig. 3: PROC curve.

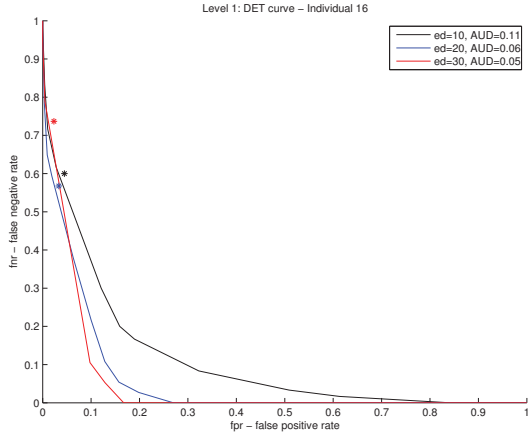

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	60	37	19
Impostor faces (total)	1646	1169	645
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1197	0.1251	0.1212
False positive rates	4.50%	3.34%	2.33%
True positive rates	40.00%	43.24%	26.32%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	4.05%	3.57%	0.00%	4.84%	11.48%	3.53%	3.95%	0.00%	8.47%	10.61%	4.05%	6.06%	2.22%	5.06%	0.00%
20 px.	2.27%	2.17%	0.00%	7.89%	6.67%	0.00%	1.64%	0.00%	5.26%	8.00%	5.77%	4.44%	1.54%	5.66%	0.00%
30 px.	0.00%	0.00%	0.00%	11.76%	4.17%	0.00%	0.00%	0.00%	10.53%	3.57%	6.67%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	**40.00%**	1.54%	0.00%	4.17%	3.77%	0.00%	0.00%	6.56%	4.94%	7.79%	0.00%	1.37%	4.84%	5.17%	1.41%
20 px.	**43.24%**	1.96%	0.00%	4.00%	2.38%	0.00%	0.00%	7.50%	3.51%	5.36%	0.00%	0.00%	2.00%	4.65%	0.00%
30 px.	**26.32%**	3.57%	0.00%	0.00%	0.00%	0.00%	0.00%	9.52%	0.00%	0.00%	0.00%	0.00%	7.14%	7.69%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 16. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

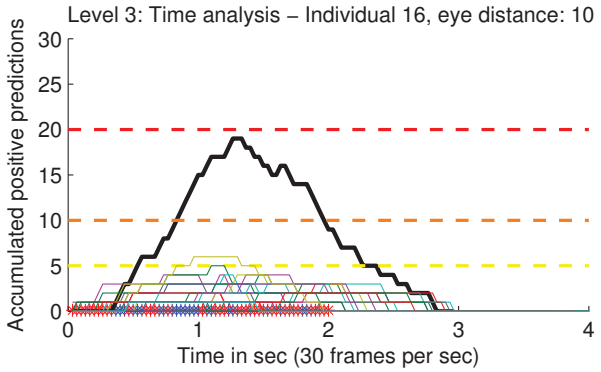

Fig. 5: Accumulated detections for 10 pixels between eyes.

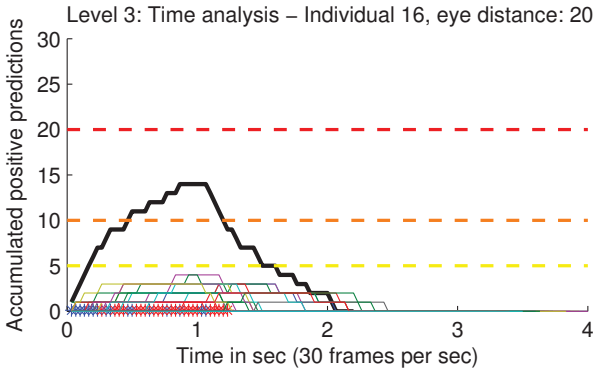

Fig. 6: Accumulated detections for 20 pixels between eyes.

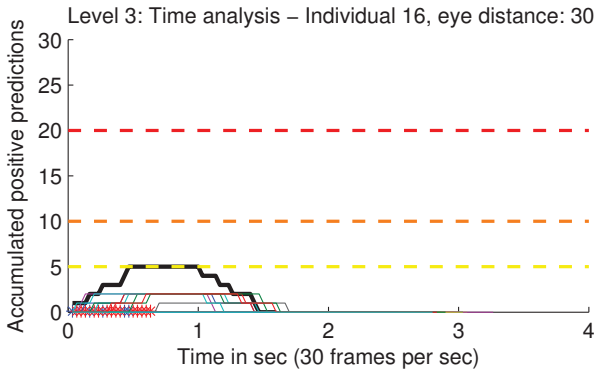

Fig. 7: Accumulated detections for 30 pixels between eyes.

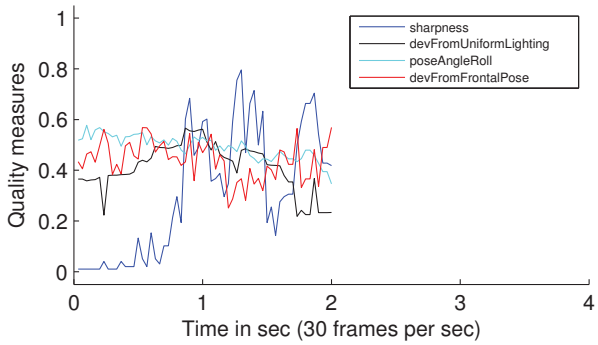

Fig. 8: Variations of quality measures.

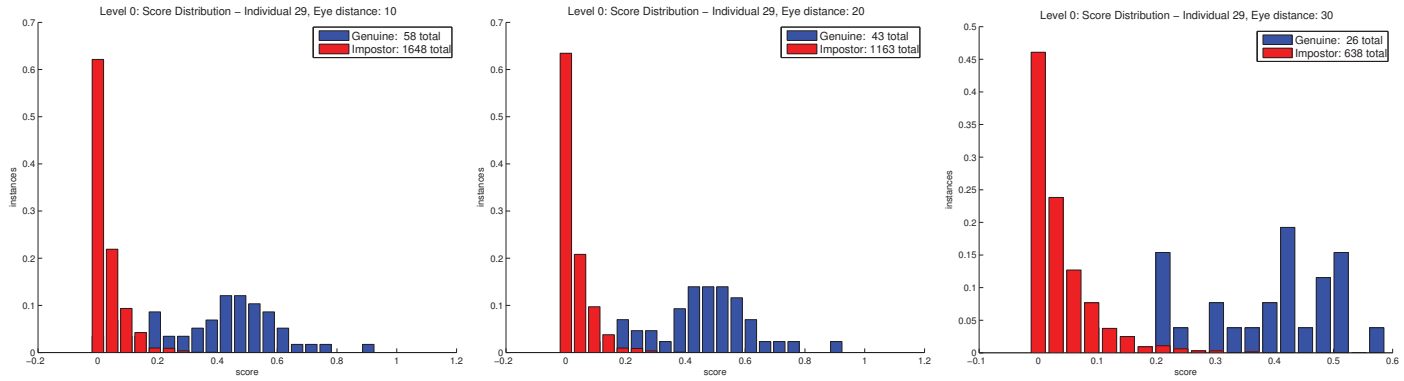

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

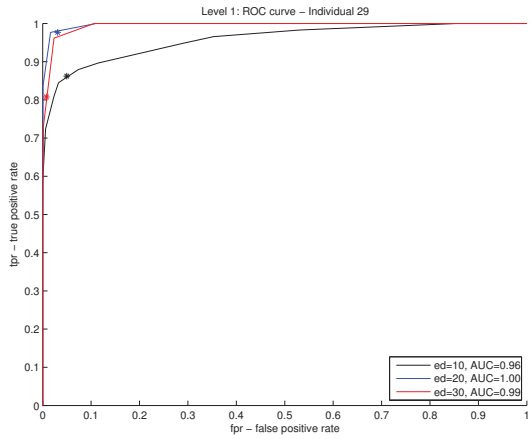

Fig. 2: ROC curve.

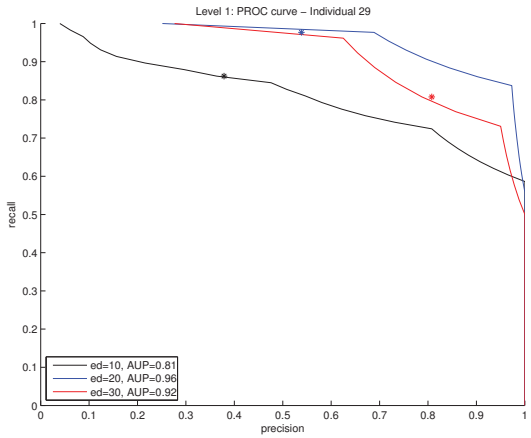

Fig. 3: PROC curve.

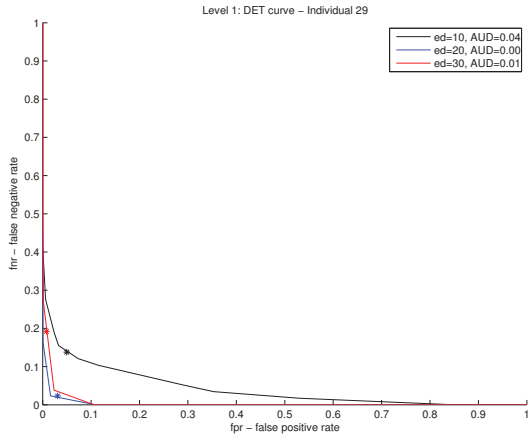

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	58	43	26
Impostor faces (total)	1648	1163	638
Detection Level			
Falsely detected faces	30.42%	11.65%	18.74%
Failure to acquire rate	2.25%	30.42%	60.96%
Matching Level			
Low quality faces	6.57%	11.72%	19.20%
Operating points	0.1569	0.1770	0.2587
False positive rates	4.98%	3.10%	0.78%
True positive rates	86.21%	97.67%	80.77%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	5.41%	0.00%	0.00%	6.45%	9.84%	1.18%	1.32%	0.00%	3.39%	4.55%	4.05%	6.06%	1.11%	3.80%	0.00%
20 px.	4.55%	0.00%	0.00%	2.63%	4.44%	0.00%	0.00%	0.00%	2.63%	2.00%	3.85%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	1.41%	4.17%	7.55%	0.00%	0.00%	1.64%	8.64%	3.90%	1.85%	4.11%	37.10%	**86.21%**	5.63%
20 px.	0.00%	0.00%	1.89%	2.00%	4.76%	0.00%	0.00%	0.00%	5.26%	3.57%	0.00%	4.08%	30.00%	**97.67%**	2.17%
30 px.	0.00%	0.00%	0.00%	0.00%	4.17%	0.00%	0.00%	0.00%	3.23%	0.00%	0.00%	0.00%	7.14%	**80.77%**	4.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 29. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

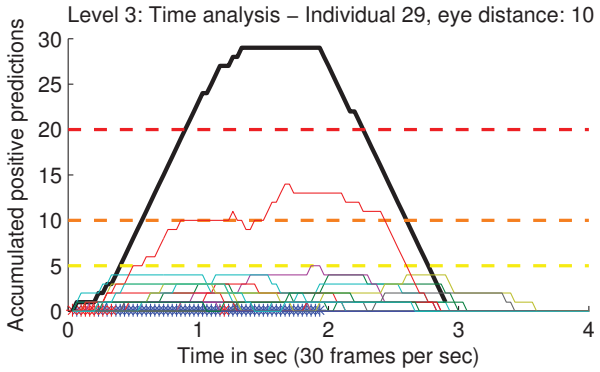

Fig. 5: Accumulated detections for 10 pixels between eyes.

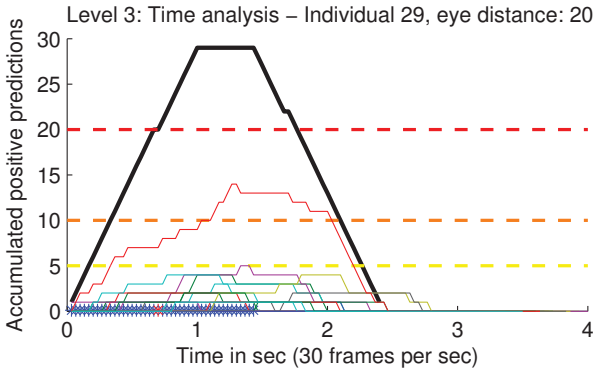

Fig. 6: Accumulated detections for 20 pixels between eyes.

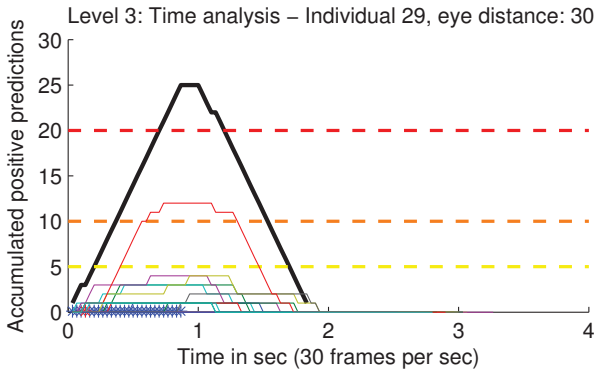

Fig. 7: Accumulated detections for 30 pixels between eyes.

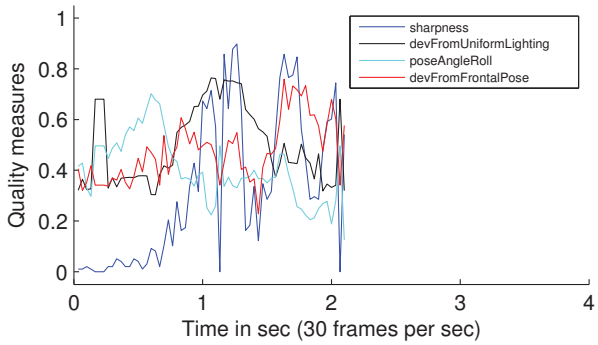

Fig. 8: Variations of quality measures.

5 Evaluation Results for PittPatt System

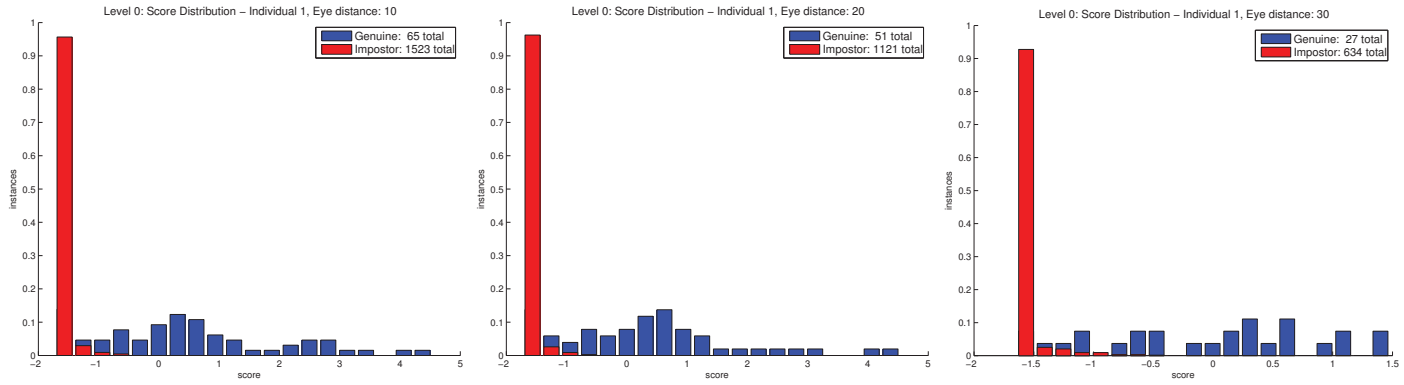

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

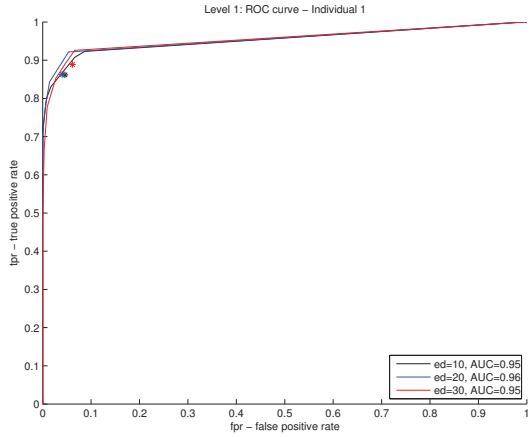

Fig. 2: ROC curve.

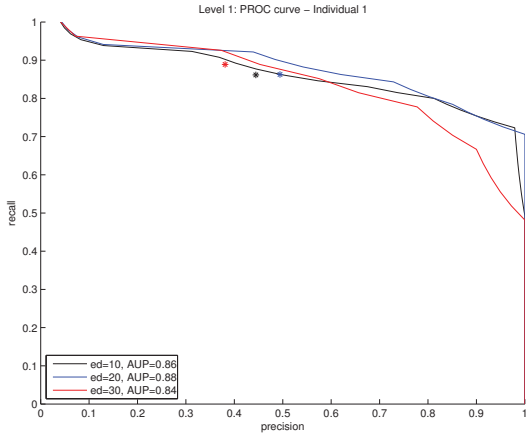

Fig. 3: PROC curve.

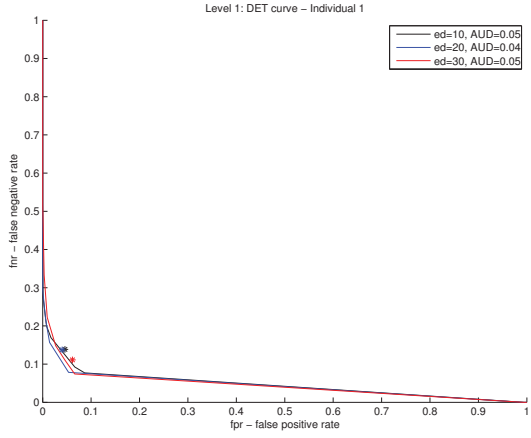

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	65	51	27
Impostor faces (total)	1523	1121	634
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.2690	-1.2726	-1.3660
False positive rates	4.60%	4.01%	6.15%
True positive rates	86.15%	86.27%	88.89%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	**86.15%**	0.00%	0.00%	0.00%	18.03%	14.10%	35.82%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	5.41%	0.00%
20 px.	**86.27%**	0.00%	0.00%	0.00%	16.67%	5.08%	42.86%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.41%	0.00%
30 px.	**88.89%**	0.00%	0.00%	0.00%	25.00%	0.00%	72.73%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	6.67%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	1.61%	1.54%	4.35%	3.64%	0.00%	0.00%	5.17%	1.28%	0.00%	0.00%	0.00%	0.00%	0.00%	14.52%
20 px.	0.00%	0.00%	1.96%	5.66%	2.44%	0.00%	0.00%	0.00%	1.72%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	3.33%	3.33%	3.45%	0.00%	0.00%	0.00%	5.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 1. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

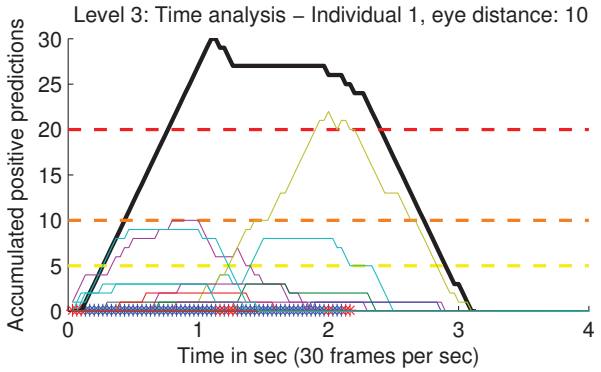

Fig. 5: Accumulated detections for 10 pixels between eyes.

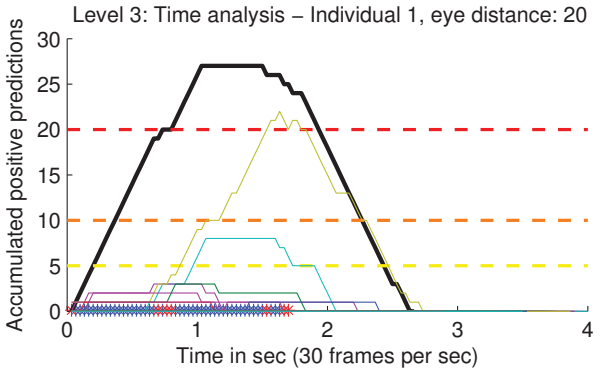

Fig. 6: Accumulated detections for 20 pixels between eyes.

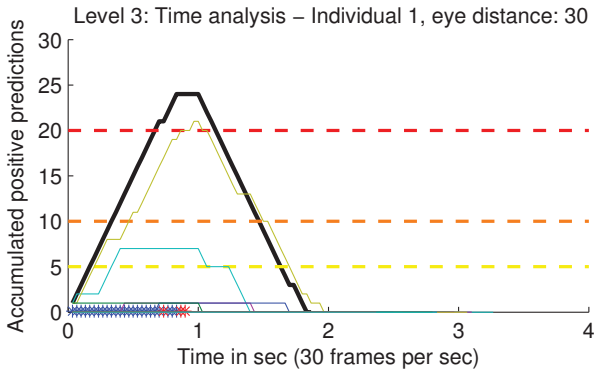

Fig. 7: Accumulated detections for 30 pixels between eyes.

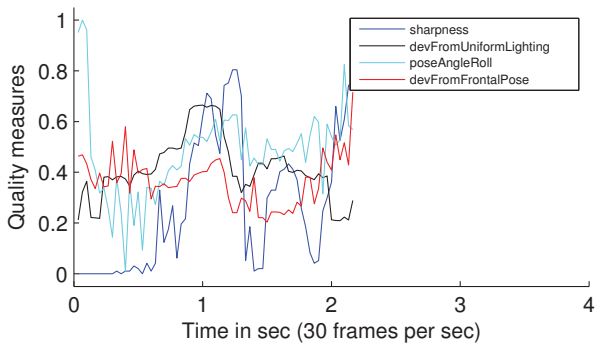

Fig. 8: Variations of quality measures.

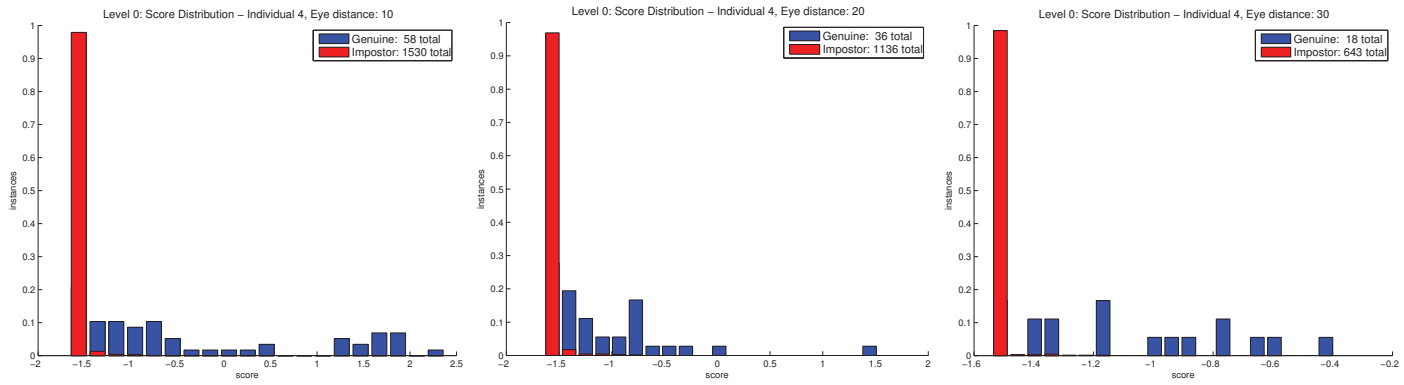

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

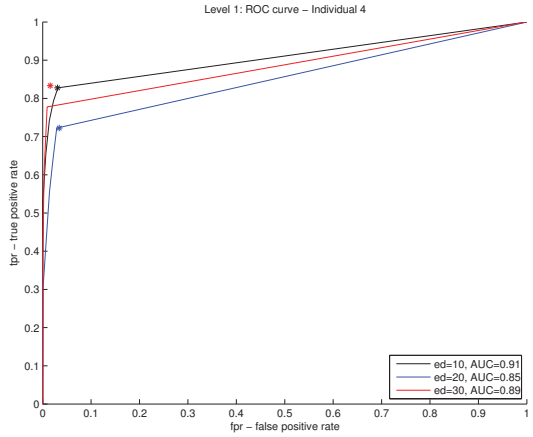

Fig. 2: ROC curve.

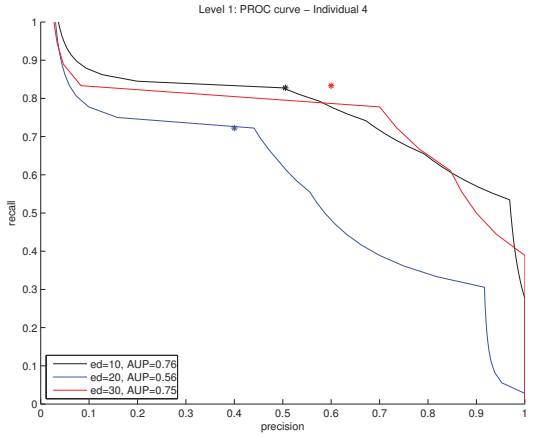

Fig. 3: PROC curve.

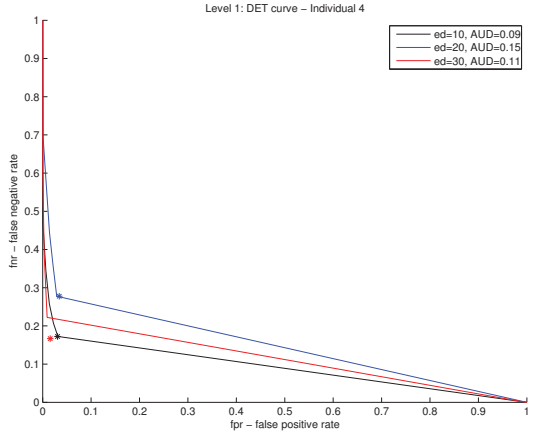

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	58	36	18
Impostor faces (total)	1530	1136	643
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.4620	-1.4515	-1.4707
False positive rates	3.07%	3.43%	1.56%
True positive rates	82.76%	72.22%	83.33%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	4.62%	0.00%	0.00%	**82.76%**	19.67%	0.00%	1.49%	0.00%	1.69%	0.00%	0.00%	3.08%	0.00%	1.35%	0.00%
20 px.	5.88%	0.00%	0.00%	**72.22%**	20.83%	0.00%	1.79%	0.00%	2.63%	0.00%	0.00%	4.44%	0.00%	1.85%	0.00%
30 px.	7.41%	0.00%	0.00%	**83.33%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	4.84%	3.08%	0.00%	0.00%	0.00%	0.00%	3.45%	0.00%	0.00%	17.65%	0.00%	1.69%	0.00%	16.13%
20 px.	0.00%	6.25%	3.92%	0.00%	0.00%	0.00%	0.00%	2.78%	0.00%	0.00%	15.15%	0.00%	2.17%	0.00%	19.57%
30 px.	0.00%	6.67%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.50%	0.00%	4.00%	0.00%	11.54%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 4. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

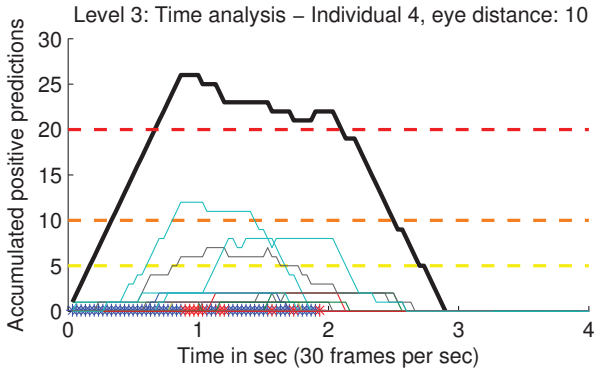

Fig. 5: Accumulated detections for 10 pixels between eyes.

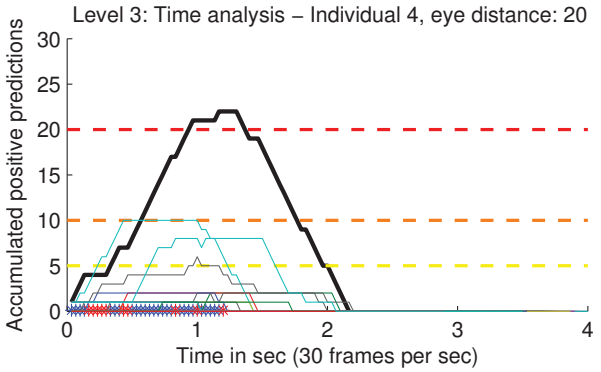

Fig. 6: Accumulated detections for 20 pixels between eyes.

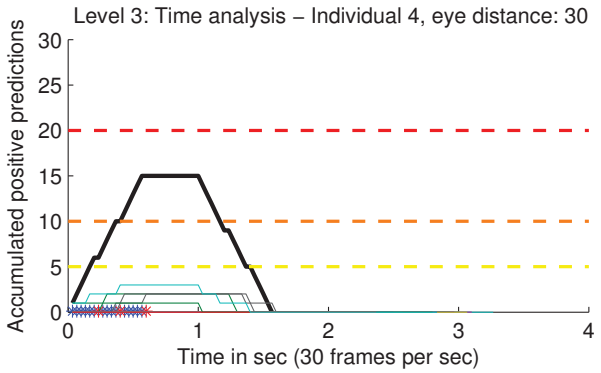

Fig. 7: Accumulated detections for 30 pixels between eyes.

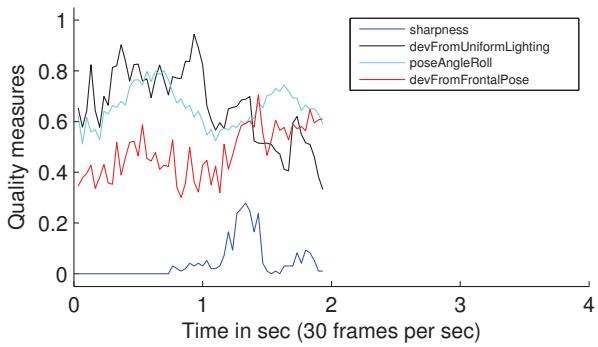

Fig. 8: Variations of quality measures.

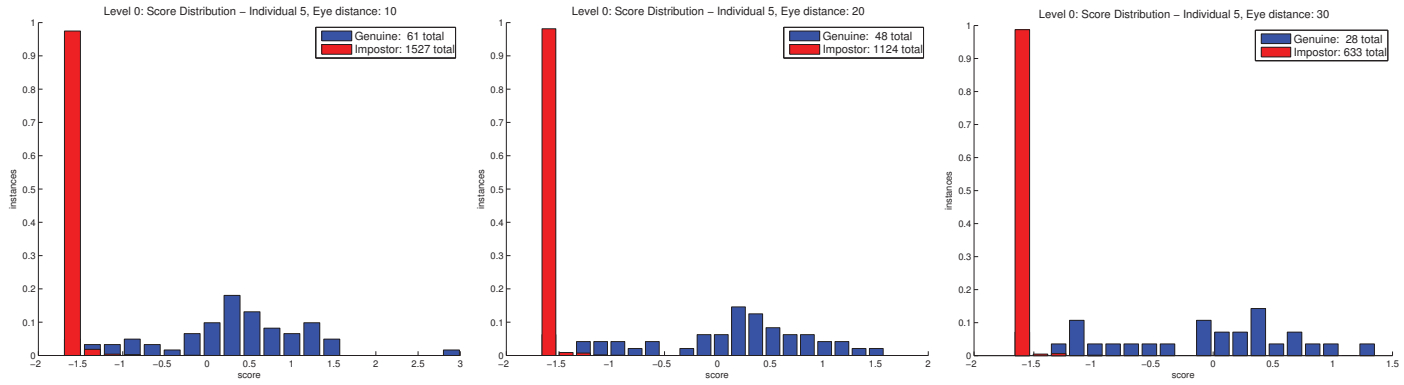

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

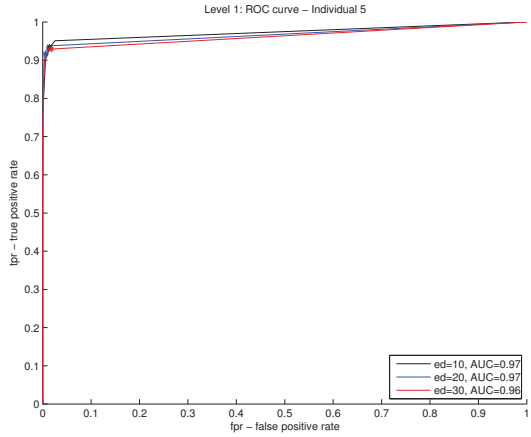

Fig. 2: ROC curve.

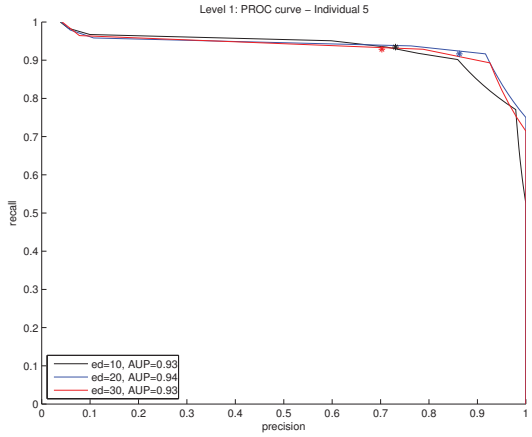

Fig. 3: PROC curve.

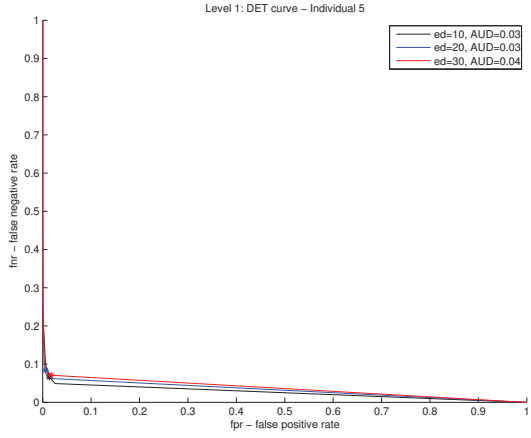

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	61	48	28
Impostor faces (total)	1527	1124	633
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.1994	-1.1964	-1.4948
False positive rates	1.38%	0.62%	1.74%
True positive rates	93.44%	91.67%	92.86%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	3.08%	0.00%	0.00%	0.00%	**93.44%**	0.00%	0.00%	0.00%	18.64%	3.08%	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	**91.67%**	0.00%	0.00%	0.00%	7.89%	4.26%	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	14.81%	0.00%	0.00%	0.00%	**92.86%**	0.00%	0.00%	0.00%	15.00%	0.00%	0.00%	0.00%	0.00%	3.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.68%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.35%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.54%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 5. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

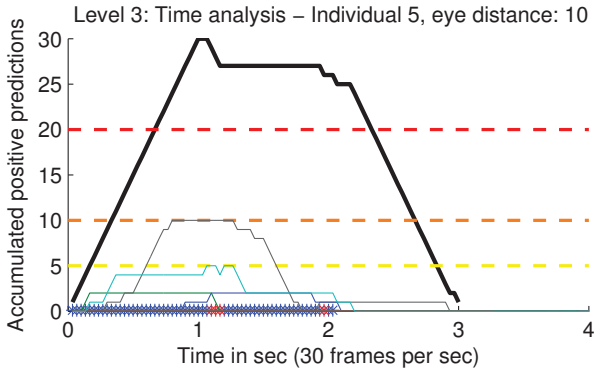

Fig. 5: Accumulated detections for 10 pixels between eyes.

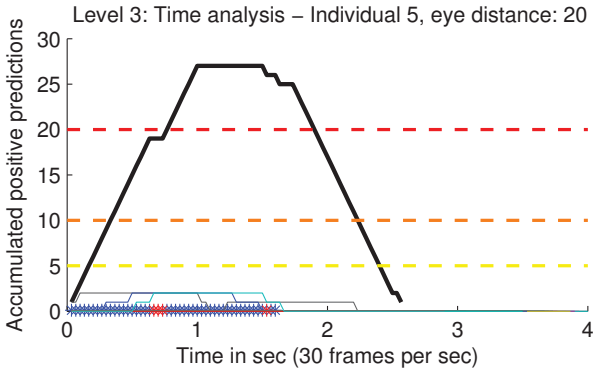

Fig. 6: Accumulated detections for 20 pixels between eyes.

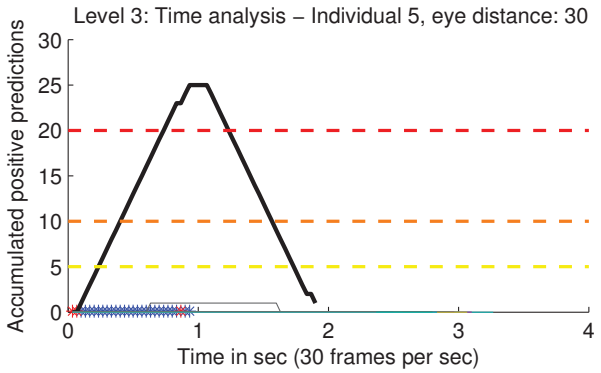

Fig. 7: Accumulated detections for 30 pixels between eyes.

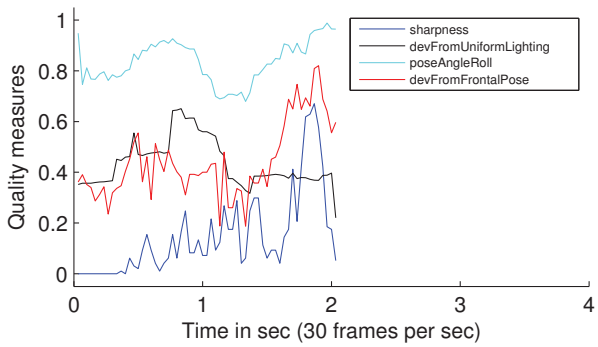

Fig. 8: Variations of quality measures.

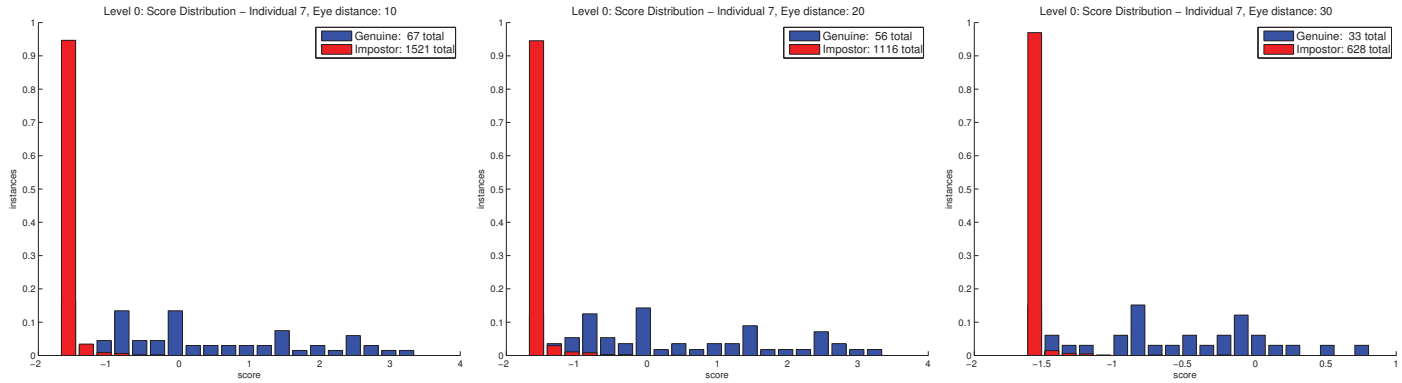

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

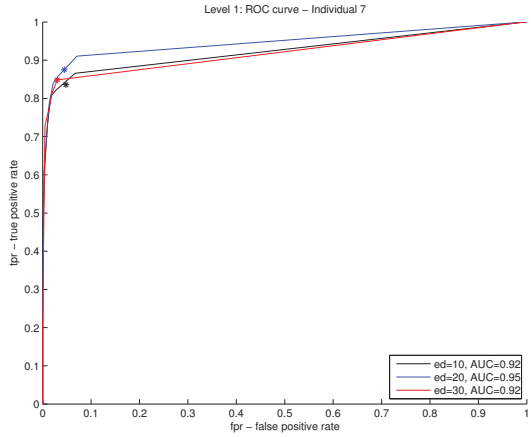

Fig. 2: ROC curve.

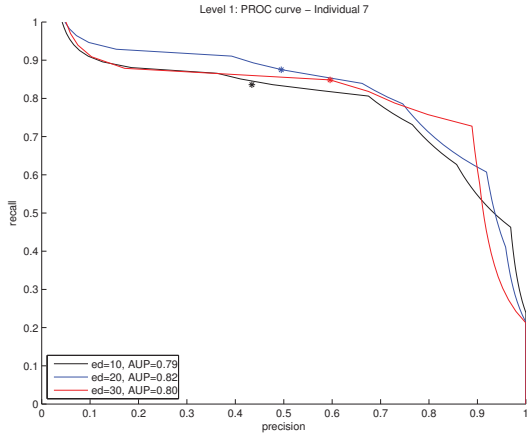

Fig. 3: PROC curve.

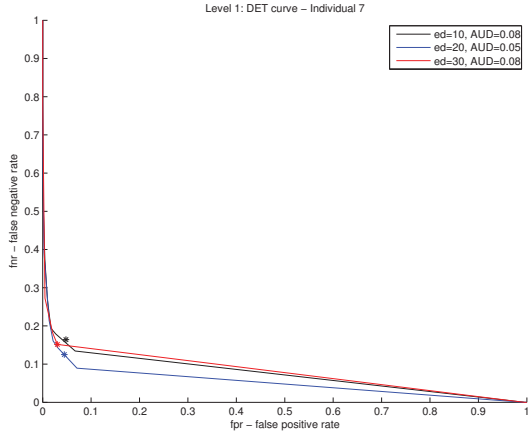

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	67	56	33
Impostor faces (total)	1521	1116	628
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.2990	-1.2595	-1.4518
False positive rates	4.80%	4.48%	3.03%
True positive rates	83.58%	87.50%	84.85%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	12.31%	0.00%	0.00%	0.00%	19.67%	0.00%	**83.58%**	0.00%	0.00%	21.54%	0.00%	0.00%	1.18%	6.76%	0.00%
20 px.	9.80%	0.00%	0.00%	0.00%	10.42%	0.00%	**87.50%**	0.00%	0.00%	23.40%	0.00%	0.00%	1.61%	7.41%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**84.85%**	0.00%	0.00%	0.00%	0.00%	0.00%	2.56%	10.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	9.68%	0.00%	0.00%	30.91%	0.00%	0.00%	8.62%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	8.06%
20 px.	0.00%	6.25%	0.00%	0.00%	36.59%	0.00%	0.00%	2.78%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	10.87%
30 px.	0.00%	13.33%	0.00%	6.90%	16.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	19.23%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 7. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

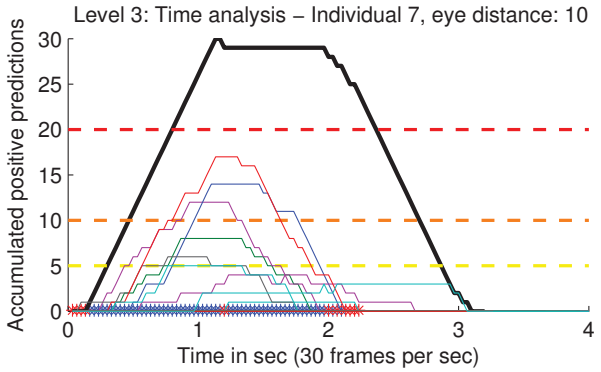

Fig. 5: Accumulated detections for 10 pixels between eyes.

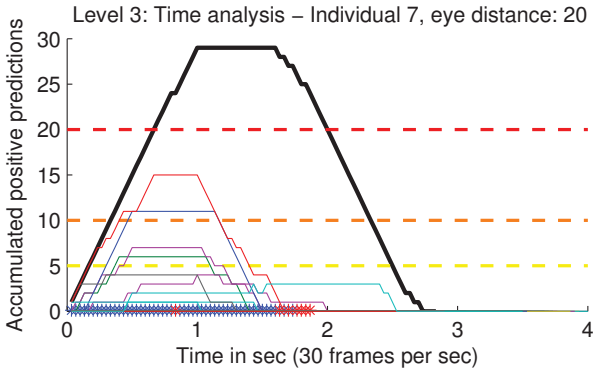

Fig. 6: Accumulated detections for 20 pixels between eyes.

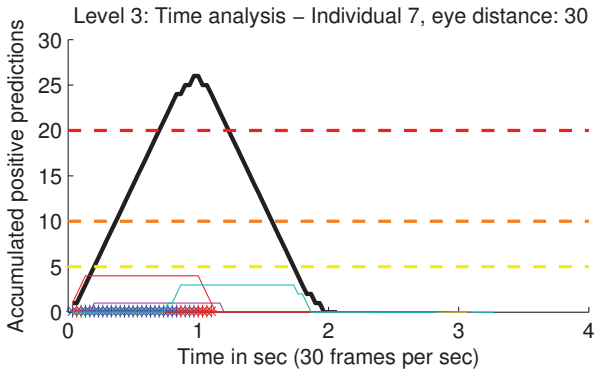

Fig. 7: Accumulated detections for 30 pixels between eyes.

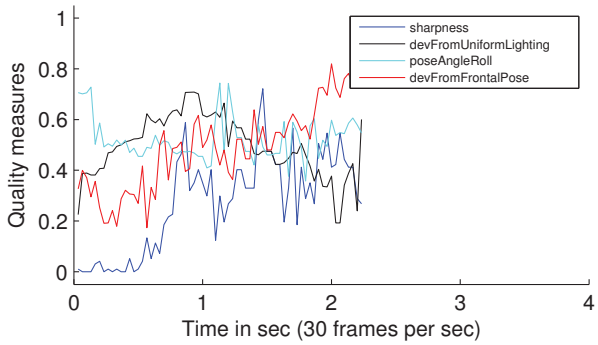

Fig. 8: Variations of quality measures.

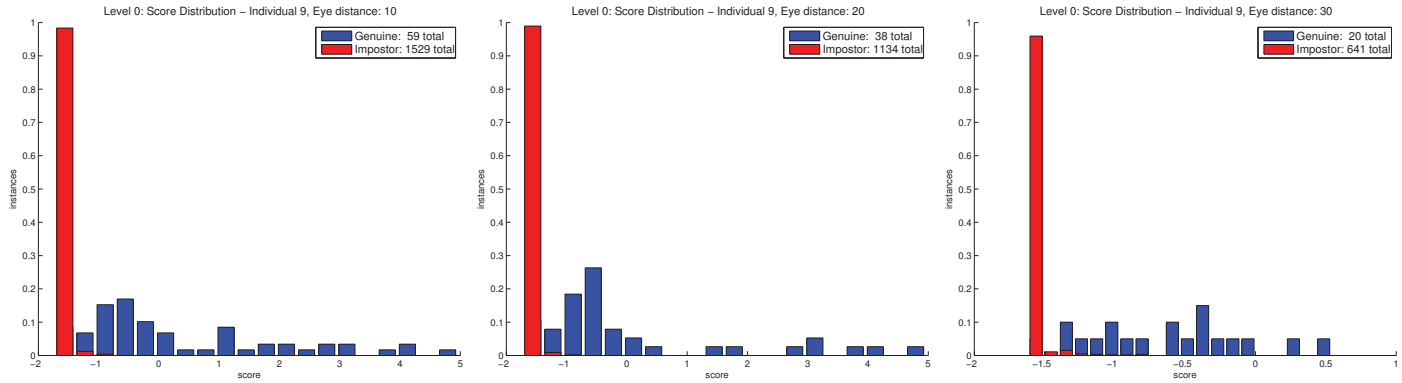

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

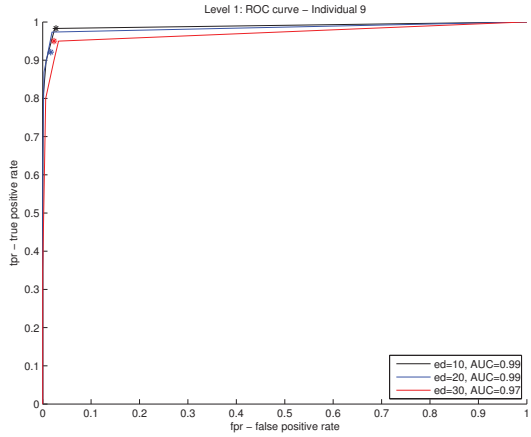

Fig. 2: ROC curve.

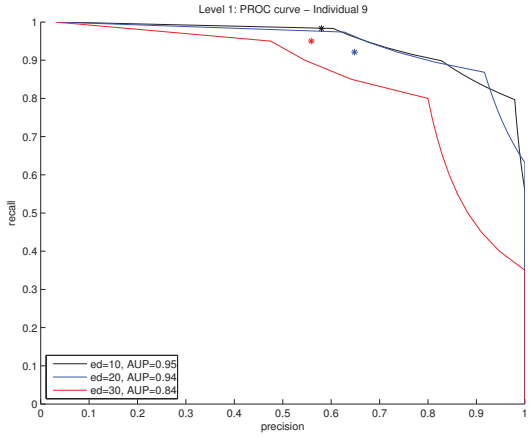

Fig. 3: PROC curve.

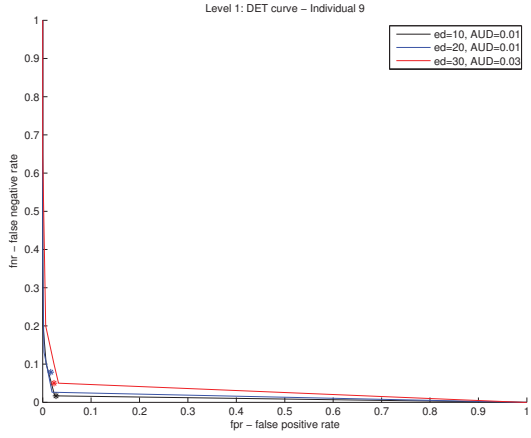

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	59	38	20
Impostor faces (total)	1529	1134	641
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.3777	-1.3172	-1.3295
False positive rates	2.75%	1.68%	2.34%
True positive rates	98.31%	92.11%	95.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	0.00%	0.00%	17.24%	8.20%	0.00%	0.00%	0.00%	**98.31%**	0.00%	0.00%	0.00%	1.18%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	8.33%	0.00%	0.00%	0.00%	**92.11%**	0.00%	0.00%	0.00%	1.61%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	10.71%	0.00%	0.00%	0.00%	**95.00%**	0.00%	0.00%	0.00%	2.56%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	1.85%	3.23%	15.38%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.17%	0.00%	0.00%	17.74%
20 px.	0.00%	0.00%	7.84%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.74%
30 px.	0.00%	0.00%	13.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	26.92%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 9. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

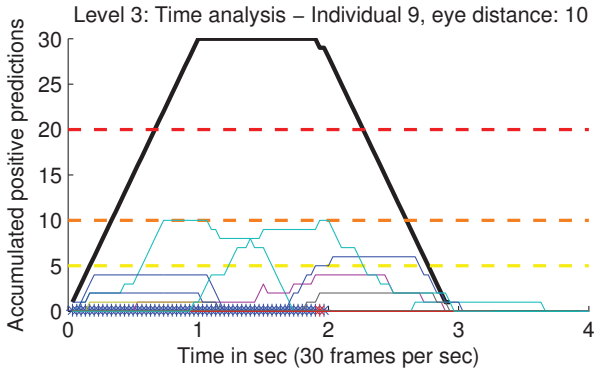

Fig. 5: Accumulated detections for 10 pixels between eyes.

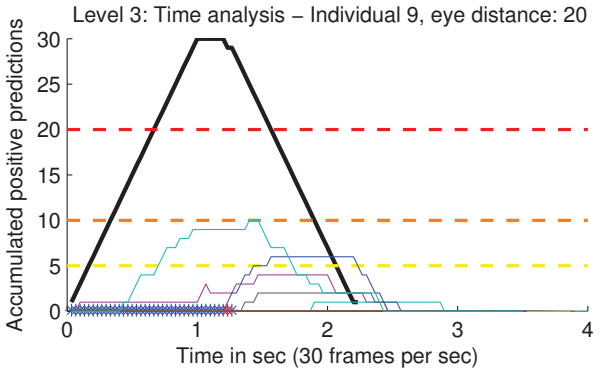

Fig. 6: Accumulated detections for 20 pixels between eyes.

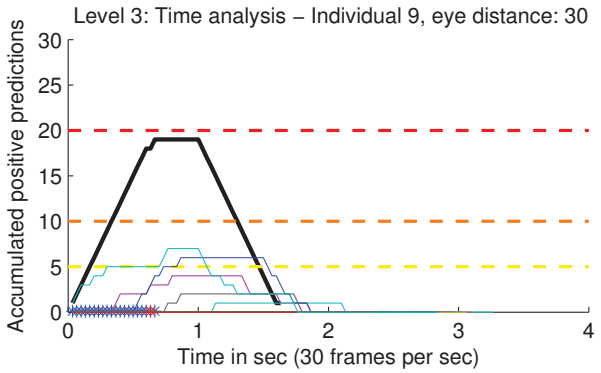

Fig. 7: Accumulated detections for 30 pixels between eyes.

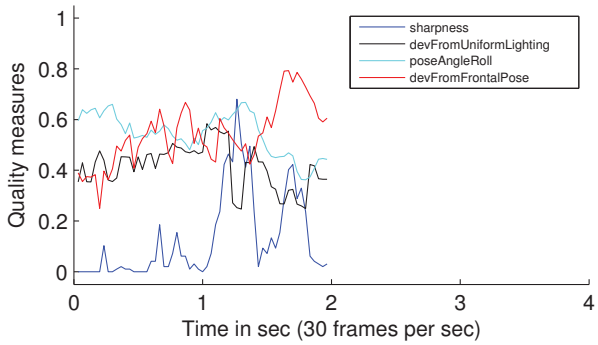

Fig. 8: Variations of quality measures.

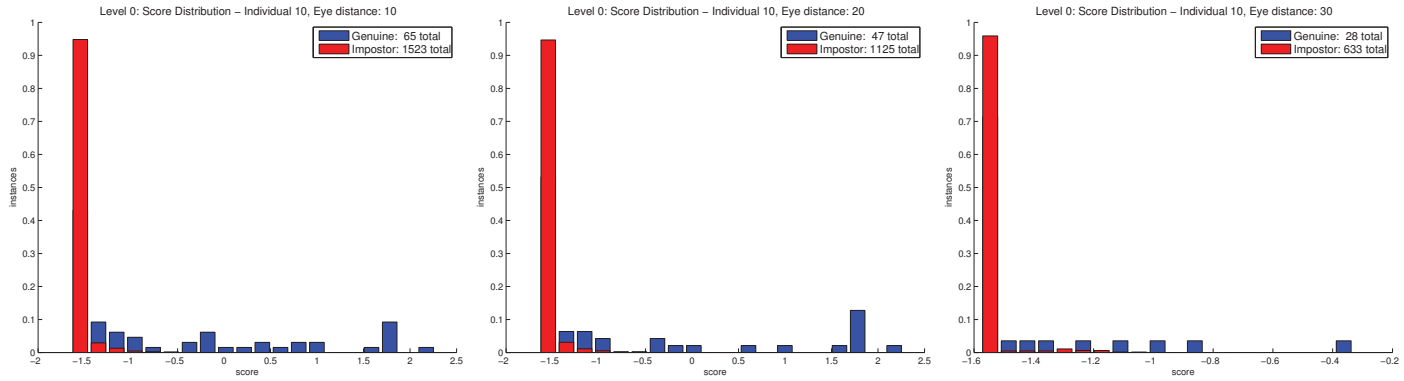

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

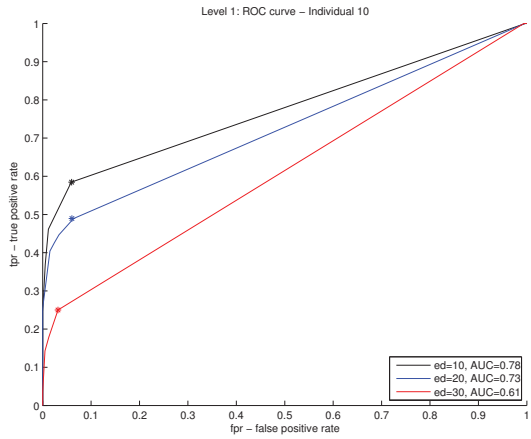

Fig. 2: ROC curve.

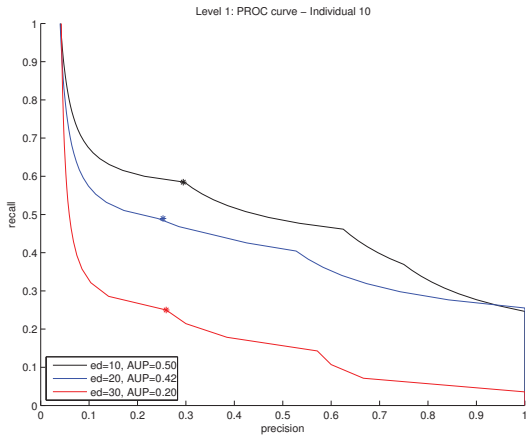

Fig. 3: PROC curve.

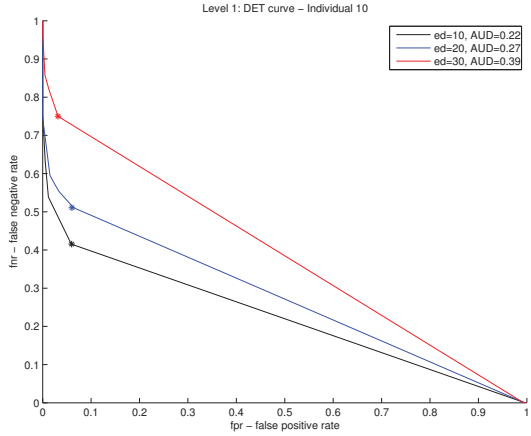

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	65	47	28
Impostor faces (total)	1523	1125	633
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.4402	-1.4121	-1.3694
False positive rates	5.98%	6.04%	3.16%
True positive rates	58.46%	48.94%	25.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	28.36%	0.00%	0.00%	**58.46%**	0.00%	0.00%	0.00%	39.19%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	33.93%	0.00%	0.00%	**48.94%**	0.00%	0.00%	0.00%	42.59%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	18.18%	0.00%	0.00%	**25.00%**	0.00%	0.00%	0.00%	33.33%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	9.68%	0.00%	1.45%	3.64%	0.00%	0.00%	12.07%	5.13%	6.67%	0.00%	0.00%	0.00%	0.00%	29.03%
20 px.	0.00%	6.25%	0.00%	1.89%	2.44%	0.00%	0.00%	2.78%	5.17%	7.02%	0.00%	0.00%	0.00%	0.00%	28.26%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	15.38%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 10. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

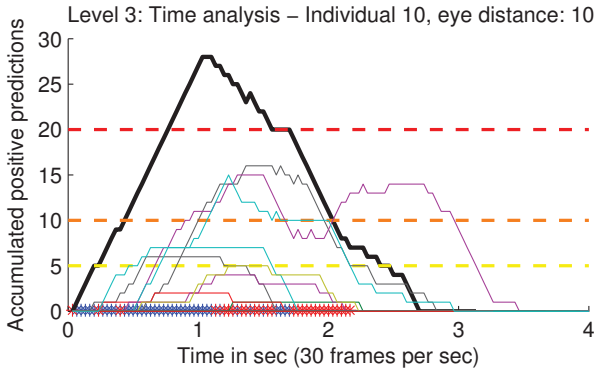

Fig. 5: Accumulated detections for 10 pixels between eyes.

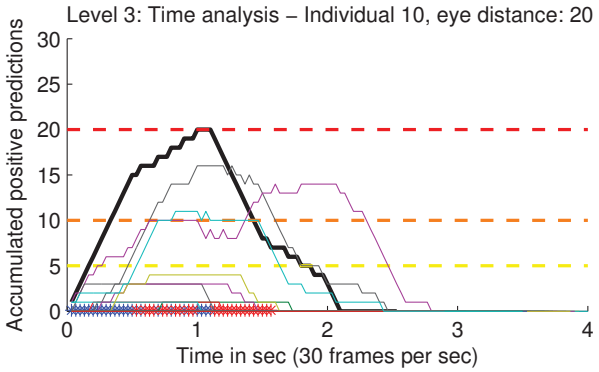

Fig. 6: Accumulated detections for 20 pixels between eyes.

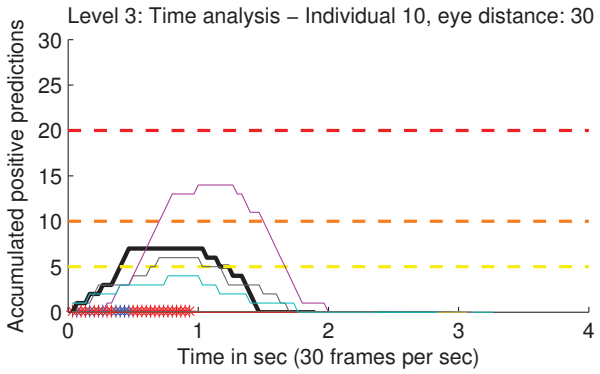

Fig. 7: Accumulated detections for 30 pixels between eyes.

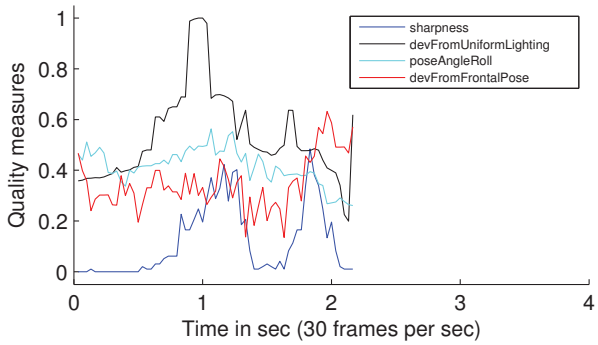

Fig. 8: Variations of quality measures.

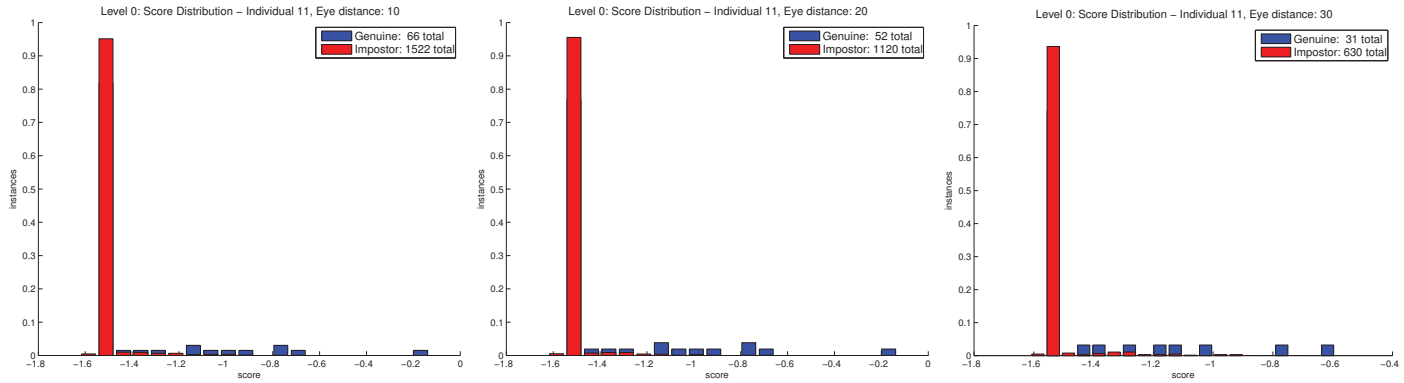

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

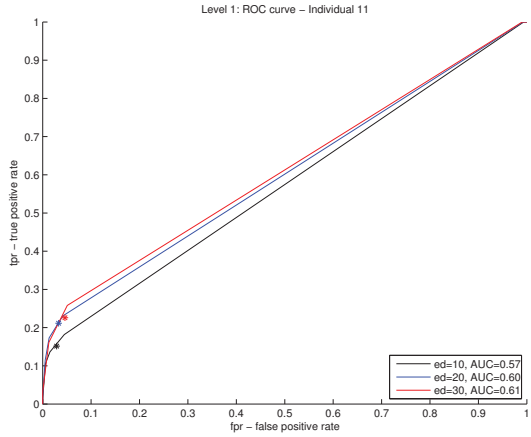

Fig. 2: ROC curve.

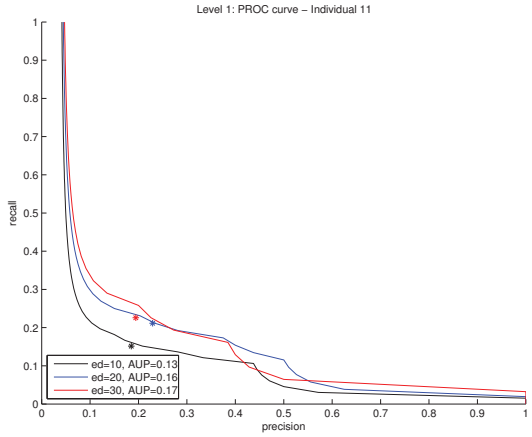

Fig. 3: PROC curve.

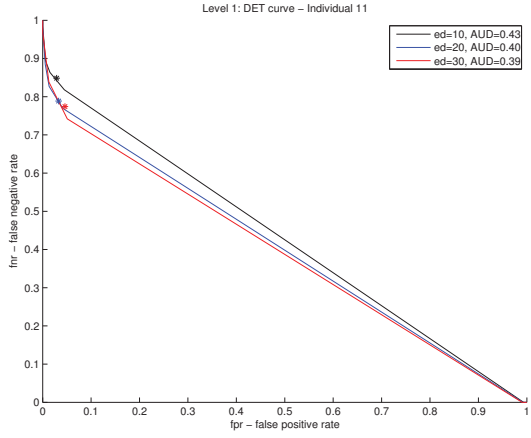

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	66	52	31
Impostor faces (total)	1522	1120	630
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.3171	-1.3667	-1.3664
False positive rates	2.89%	3.30%	4.60%
True positive rates	15.15%	21.15%	22.58%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	0.00%	0.00%	0.00%	4.92%	1.28%	0.00%	0.00%	0.00%	0.00%	**15.15%**	0.00%	9.41%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	6.25%	1.69%	0.00%	0.00%	0.00%	0.00%	**21.15%**	0.00%	12.90%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	10.71%	2.94%	0.00%	0.00%	0.00%	0.00%	**22.58%**	0.00%	20.51%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	3.70%	0.00%	0.00%	0.00%	3.64%	0.00%	0.00%	3.45%	7.69%	6.67%	0.00%	0.00%	0.00%	0.00%	24.19%
20 px.	5.26%	0.00%	0.00%	0.00%	2.44%	0.00%	0.00%	5.56%	10.34%	3.51%	0.00%	0.00%	0.00%	0.00%	26.09%
30 px.	10.53%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.11%	2.94%	6.06%	0.00%	0.00%	0.00%	0.00%	38.46%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 11. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

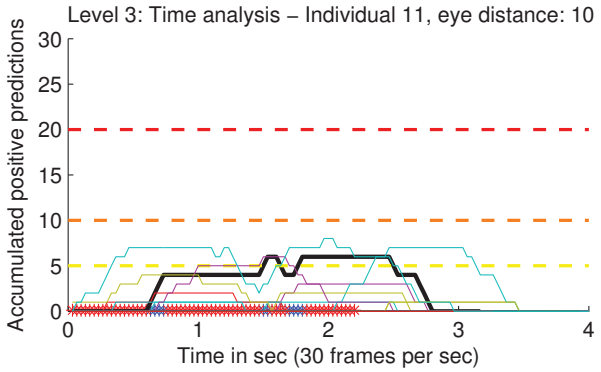

Fig. 5: Accumulated detections for 10 pixels between eyes.

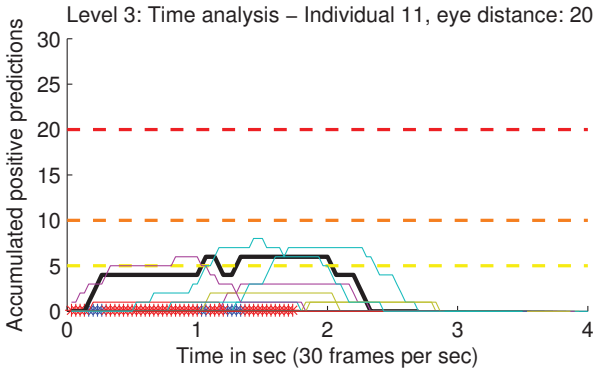

Fig. 6: Accumulated detections for 20 pixels between eyes.

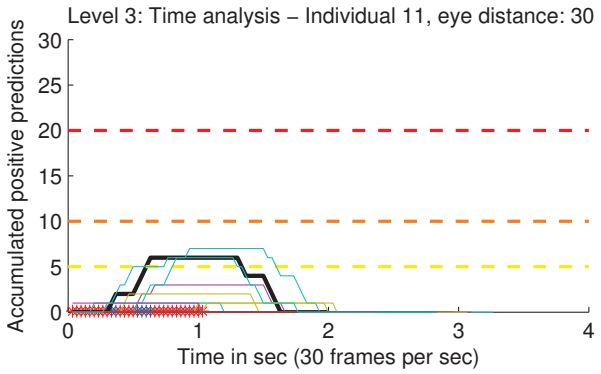

Fig. 7: Accumulated detections for 30 pixels between eyes.

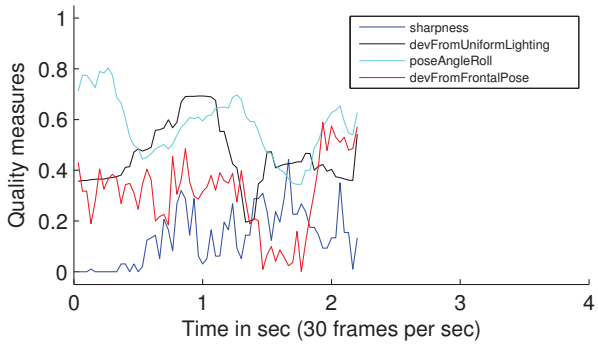

Fig. 8: Variations of quality measures.

Biometric System: PittPatt	**Biometric Modality**: Face
Data Source: Chokepoint – Portal 2, leaving portal setup, session 4, camera 1.1	**Individual Template**: 12

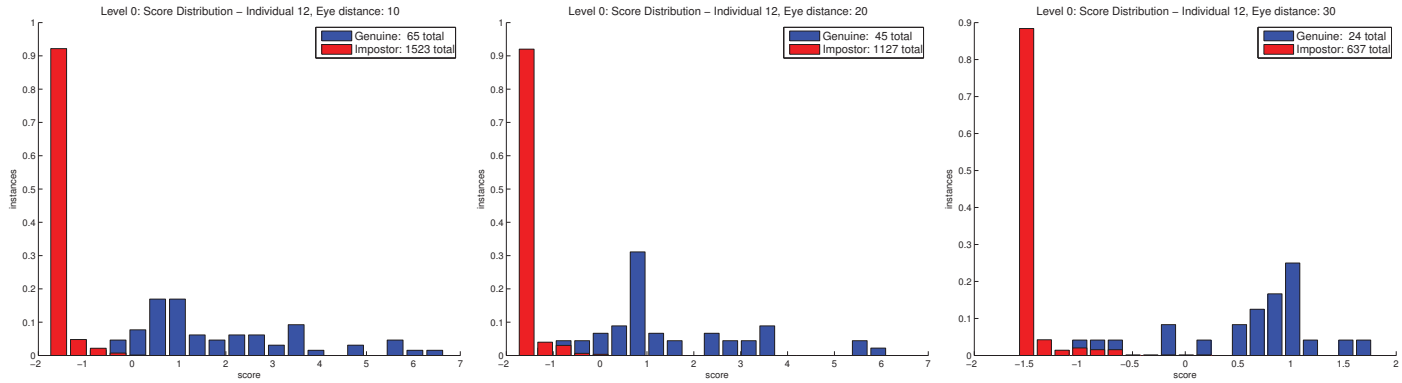

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

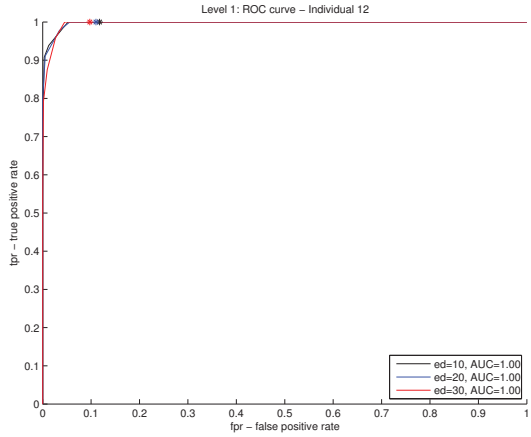

Fig. 2: ROC curve.

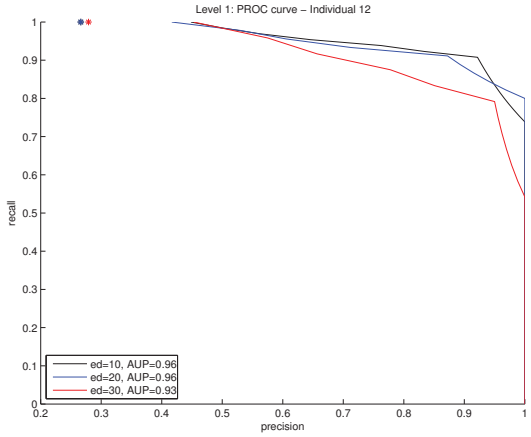

Fig. 3: PROC curve.

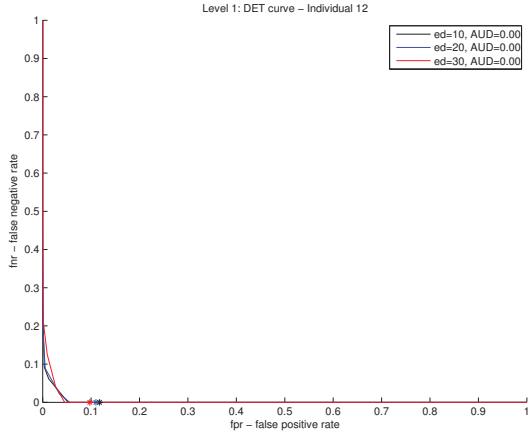

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	65	45	24
Impostor faces (total)	1523	1127	637
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.3507	-1.3246	-1.2752
False positive rates	11.75%	11.00%	9.73%
True positive rates	100.00%	100.00%	100.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	24.62%	0.00%	0.00%	29.31%	26.23%	0.00%	0.00%	0.00%	47.46%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%
20 px.	29.41%	0.00%	0.00%	19.44%	27.08%	0.00%	0.00%	0.00%	36.84%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%
30 px.	48.15%	0.00%	0.00%	0.00%	39.29%	0.00%	0.00%	0.00%	10.00%	0.00%	0.00%	**100.00%**	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	64.62%	0.00%	1.82%	0.00%	0.00%	0.00%	0.00%	0.00%	92.16%	6.35%	0.00%	0.00%	12.90%
20 px.	0.00%	0.00%	64.71%	0.00%	2.44%	0.00%	0.00%	0.00%	0.00%	0.00%	90.91%	6.38%	0.00%	0.00%	17.39%
30 px.	0.00%	0.00%	53.33%	0.00%	4.00%	0.00%	0.00%	0.00%	0.00%	0.00%	87.50%	11.54%	0.00%	0.00%	7.69%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 12. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

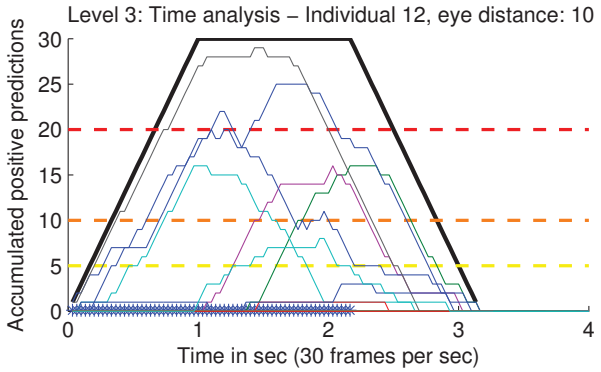

Fig. 5: Accumulated detections for 10 pixels between eyes.

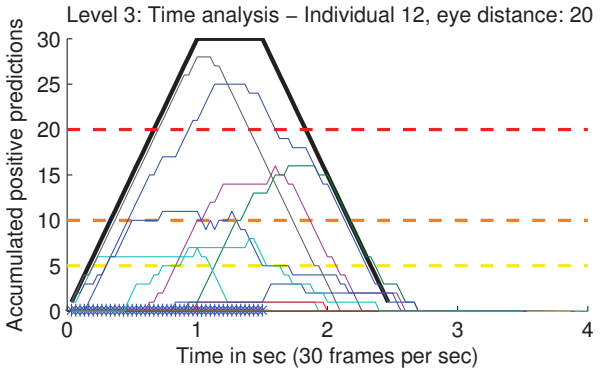

Fig. 6: Accumulated detections for 20 pixels between eyes.

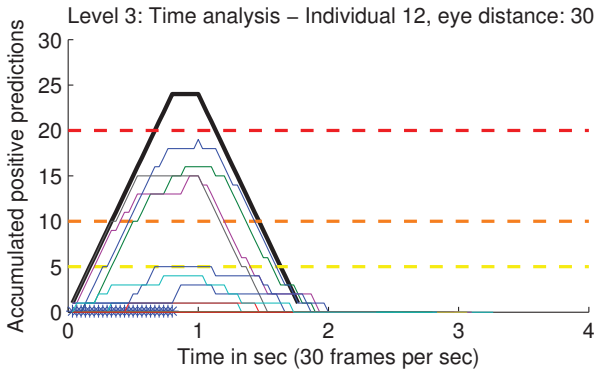

Fig. 7: Accumulated detections for 30 pixels between eyes.

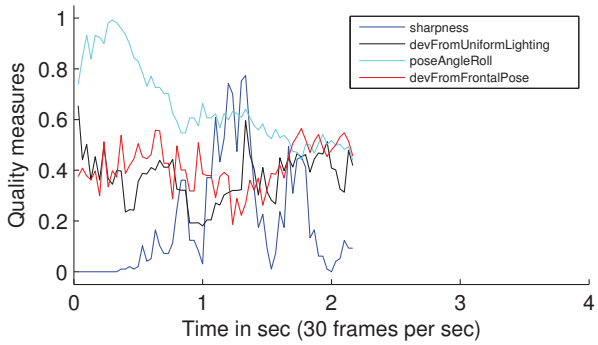

Fig. 8: Variations of quality measures.

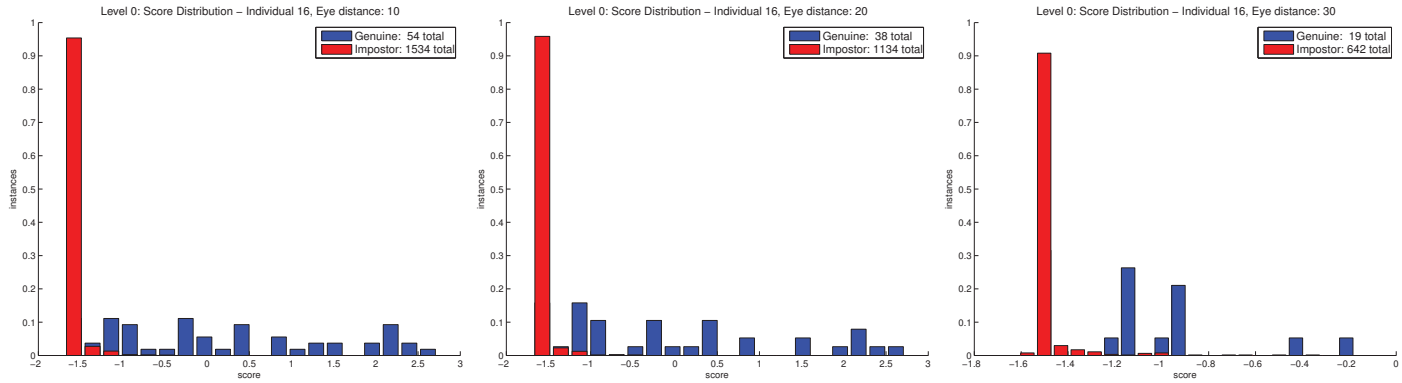

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

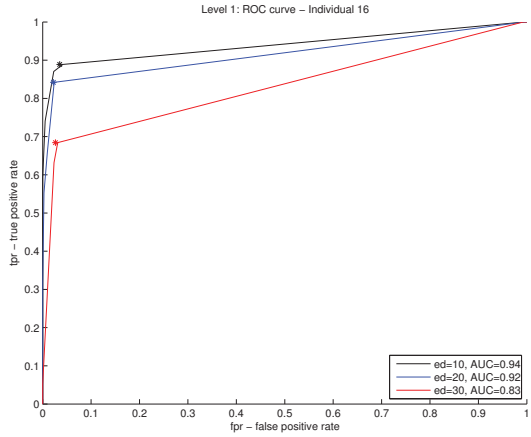

Fig. 2: ROC curve.

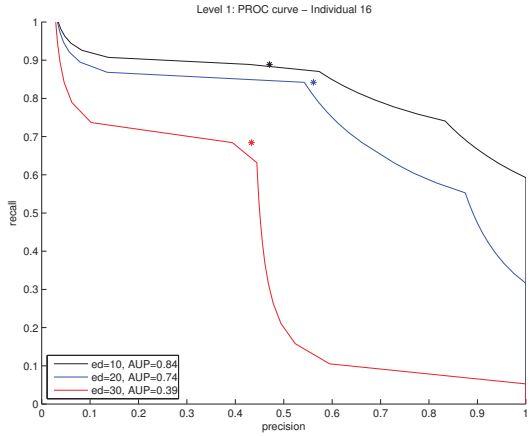

Fig. 3: PROC curve.

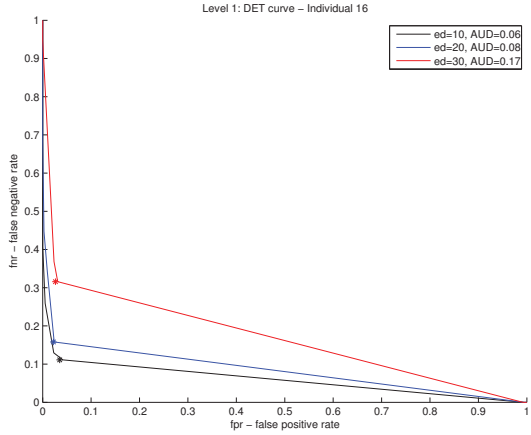

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	54	38	19
Impostor faces (total)	1534	1134	642
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	-1.2928	-1.2210	-1.2151
False positive rates	3.52%	2.20%	2.65%
True positive rates	88.89%	84.21%	68.42%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	12.31%	0.00%	0.00%	5.17%	9.84%	1.28%	0.00%	0.00%	3.39%	13.85%	0.00%	0.00%	0.00%	8.11%	0.00%
20 px.	13.73%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.63%	4.26%	0.00%	0.00%	0.00%	11.11%	0.00%
30 px.	25.93%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.57%	0.00%	0.00%	0.00%	20.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	**88.89%**	1.61%	0.00%	0.00%	0.00%	0.00%	0.00%	1.72%	0.00%	5.33%	0.00%	0.00%	0.00%	0.00%	20.97%
20 px.	**84.21%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	19.57%
30 px.	**68.42%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.54%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 16. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

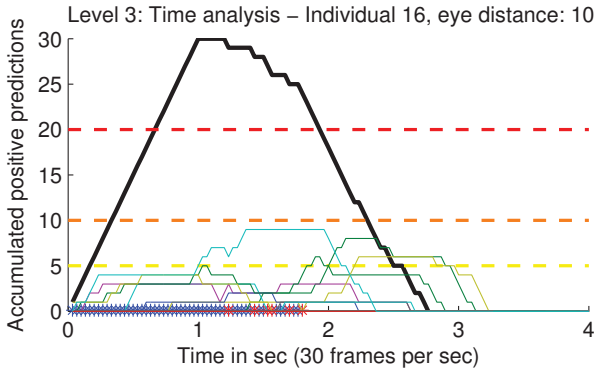

Fig. 5: Accumulated detections for 10 pixels between eyes.

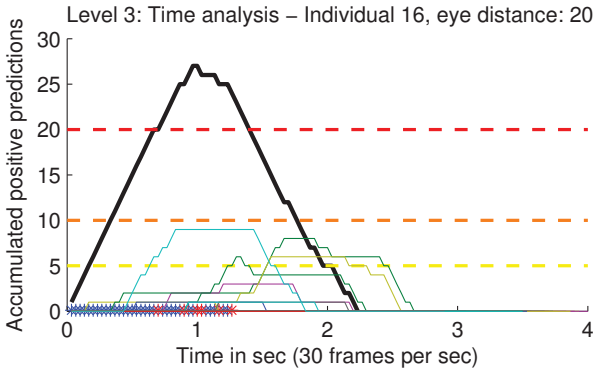

Fig. 6: Accumulated detections for 20 pixels between eyes.

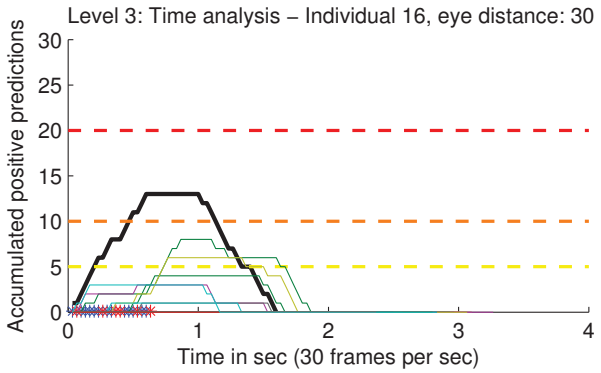

Fig. 7: Accumulated detections for 30 pixels between eyes.

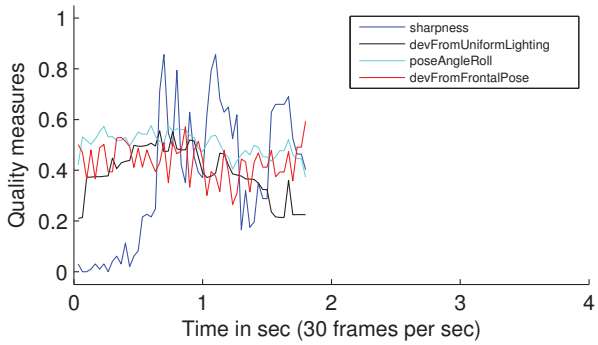

Fig. 8: Variations of quality measures.

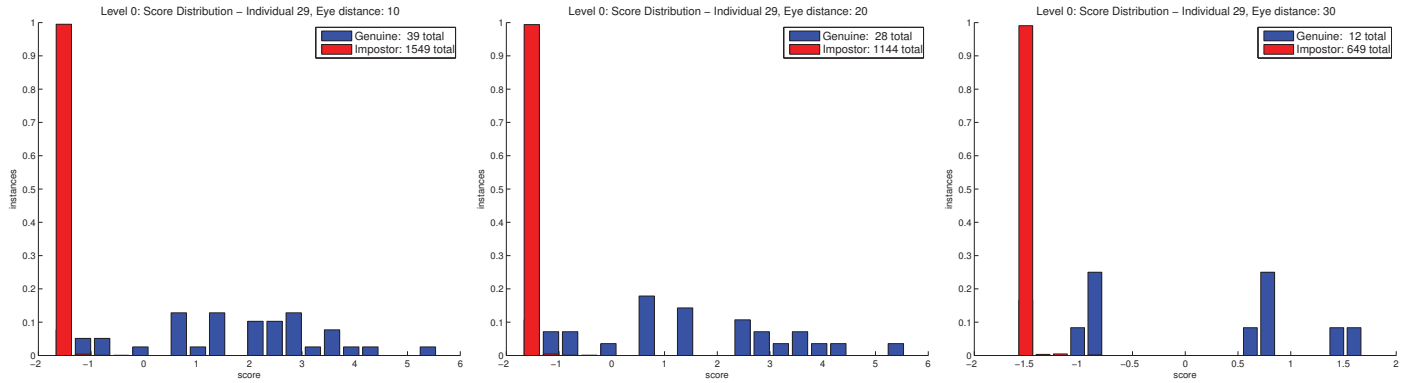

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

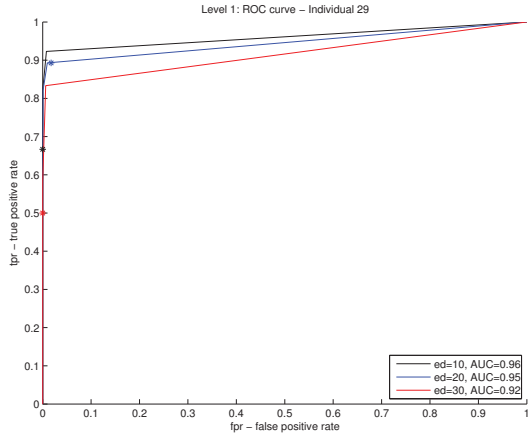

Fig. 2: ROC curve.

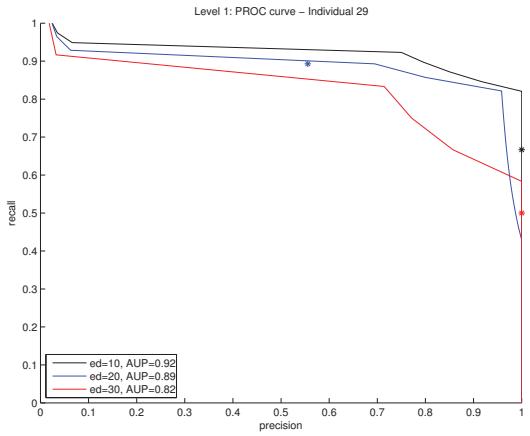

Fig. 3: PROC curve.

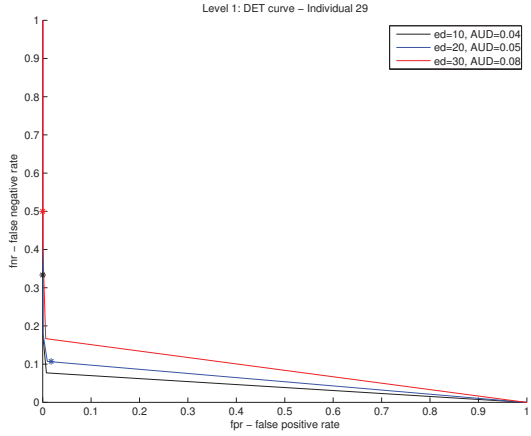

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	39	28	12
Impostor faces (total)	1549	1144	649
Detection Level			
Falsely detected faces	1.79%	1.10%	1.93%
Failure to acquire rate	10.54%	33.97%	62.76%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	0.9858	-1.4156	-0.1539
False positive rates	0.00%	1.75%	0.00%
True positive rates	66.67%	89.29%	50.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**66.67%**	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	3.03%	0.00%	28.26%	**89.29%**	4.35%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	**50.00%**	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 29. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

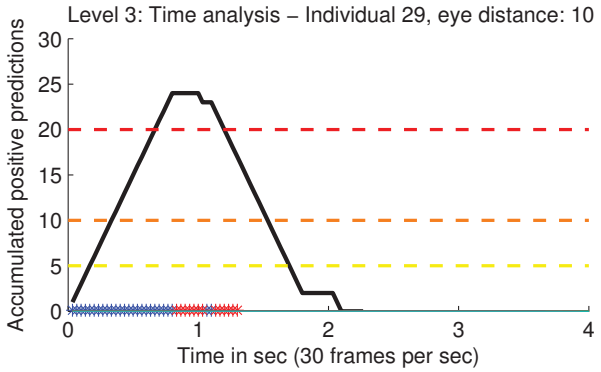

Fig. 5: Accumulated detections for 10 pixels between eyes.

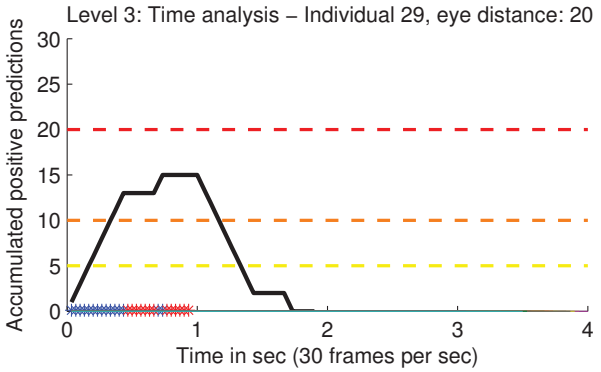

Fig. 6: Accumulated detections for 20 pixels between eyes.

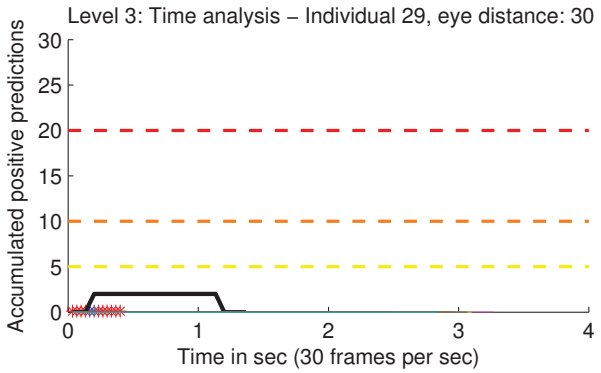

Fig. 7: Accumulated detections for 30 pixels between eyes.

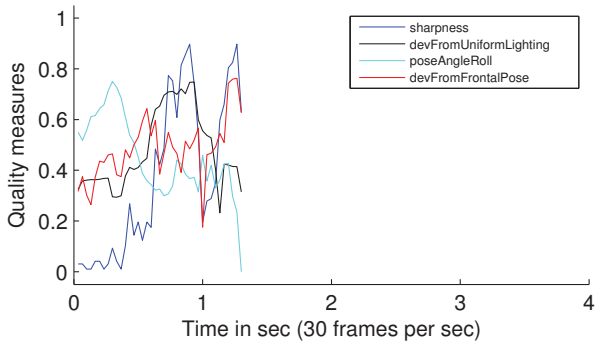

Fig. 8: Variations of quality measures.

6 Evaluation Results for Neurotechnology System

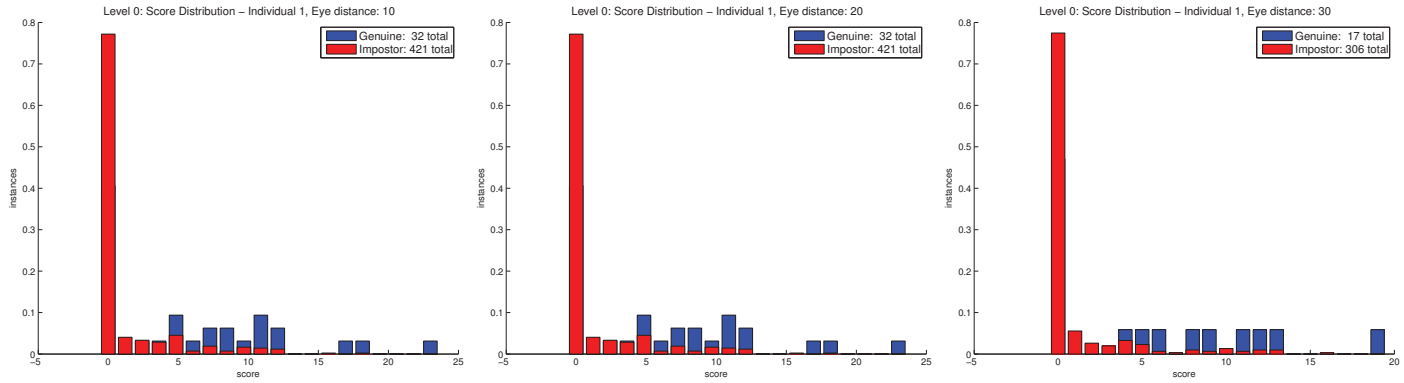

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

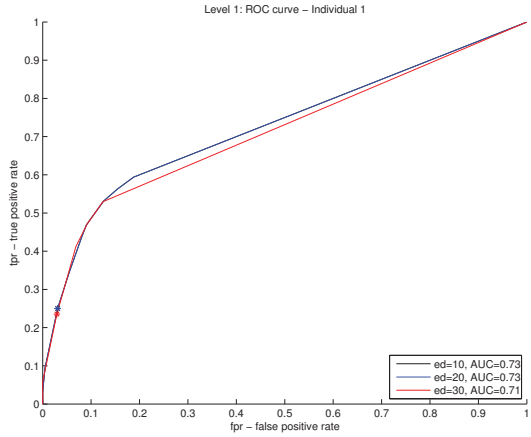

Fig. 2: ROC curve.

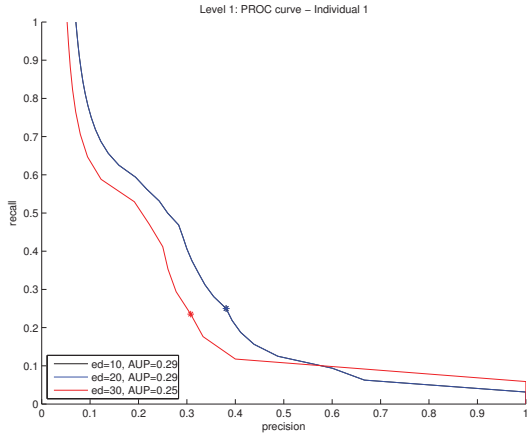

Fig. 3: PROC curve.

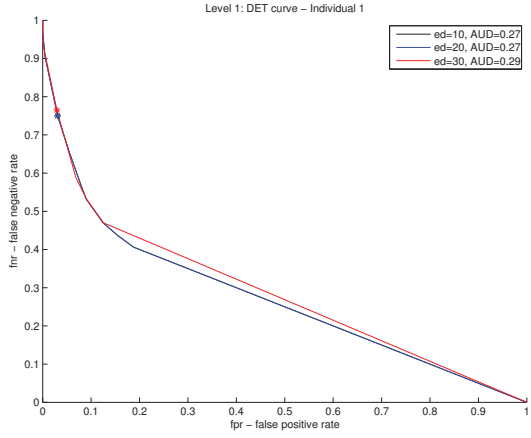

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	32	32	17
Impostor faces (total)	421	421	306
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	10.7913	10.7913	10.1951
False positive rates	3.09%	3.09%	2.94%
True positive rates	25.00%	25.00%	23.53%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	**25.00%**	0.00%	0.00%	11.11%	0.00%	0.00%	4.55%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	**25.00%**	0.00%	0.00%	11.11%	0.00%	0.00%	4.55%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	**23.53%**	0.00%	0.00%	11.11%	0.00%	0.00%	5.88%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	0.00%	0.00%	0.00%	26.67%	8.70%	0.00%	33.33%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	0.00%	0.00%	0.00%	26.67%	8.70%	0.00%	33.33%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	0.00%	0.00%	0.00%	26.67%	7.69%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 1. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

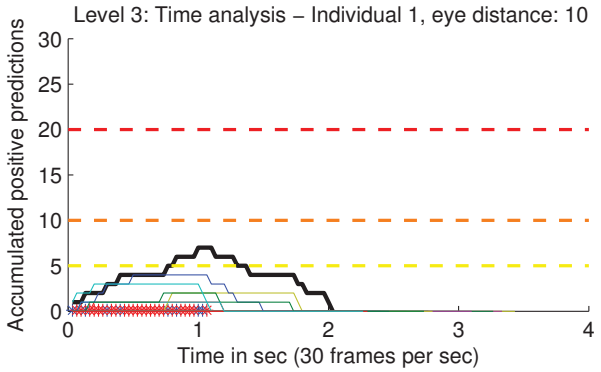

Fig. 5: Accumulated detections for 10 pixels between eyes.

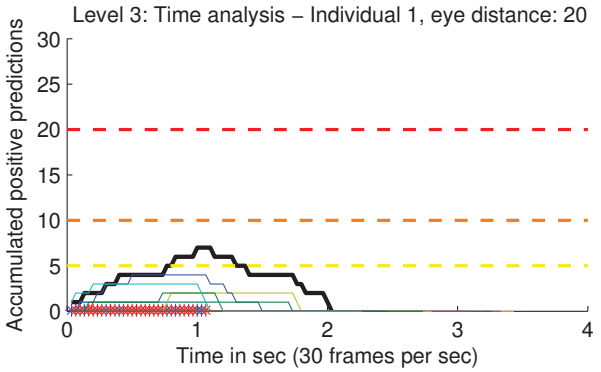

Fig. 6: Accumulated detections for 20 pixels between eyes.

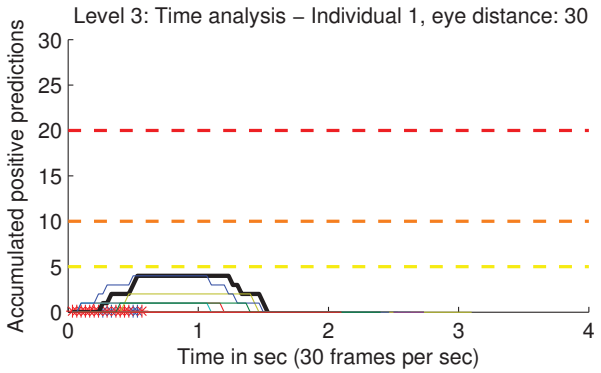

Fig. 7: Accumulated detections for 30 pixels between eyes.

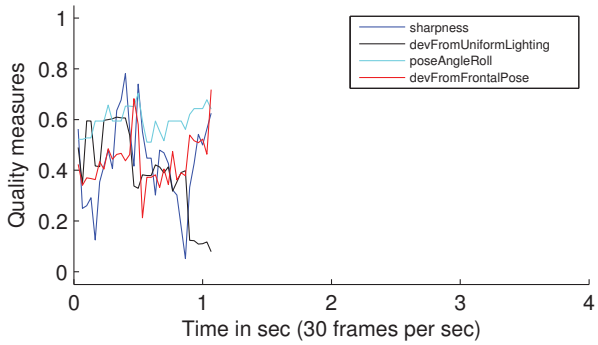

Fig. 8: Variations of quality measures.

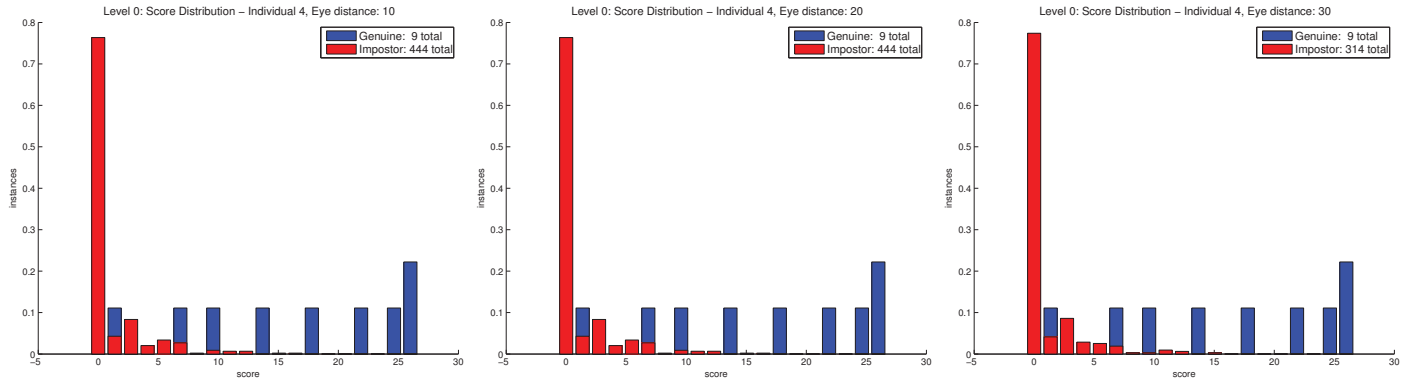

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

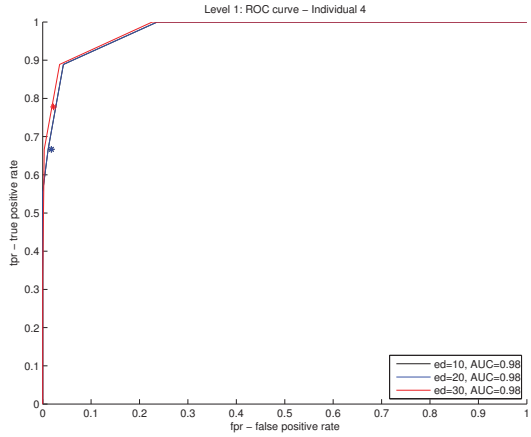

Fig. 2: ROC curve.

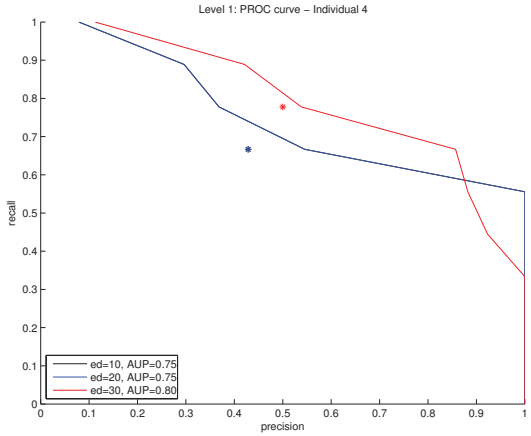

Fig. 3: PROC curve.

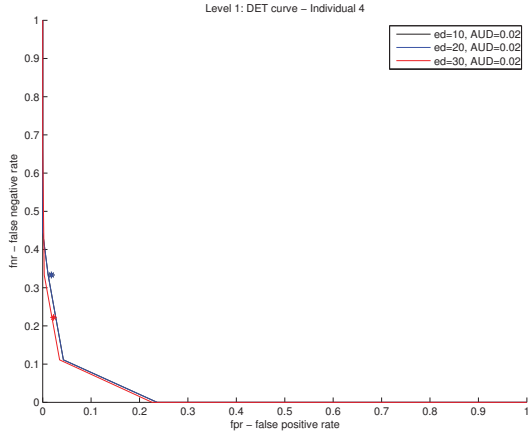

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	9	9	9
Impostor faces (total)	444	444	314
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	10.1895	10.1895	9.7923
False positive rates	1.80%	1.80%	2.23%
True positive rates	66.67%	66.67%	77.78%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	0.00%	0.00%	0.00%	**66.67%**	0.00%	10.53%	2.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	**66.67%**	0.00%	10.53%	2.27%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	**77.78%**	0.00%	11.76%	2.94%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	6.67%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.33%	4.35%	0.00%	0.00%
20 px.	6.67%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.33%	4.35%	0.00%	0.00%
30 px.	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	20.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 4. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

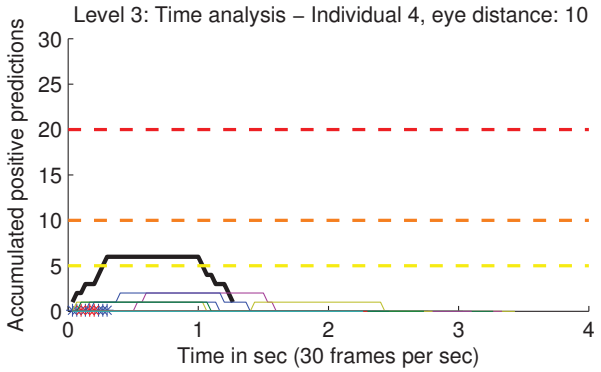

Fig. 5: Accumulated detections for 10 pixels between eyes.

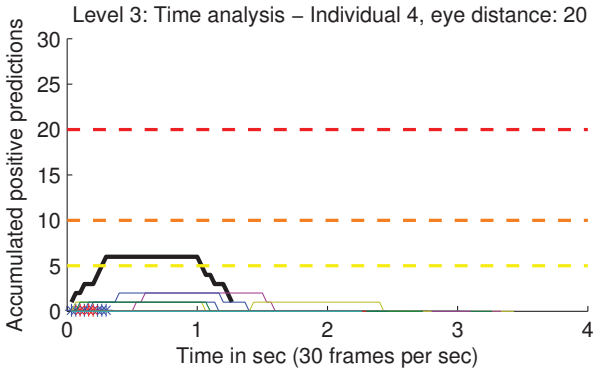

Fig. 6: Accumulated detections for 20 pixels between eyes.

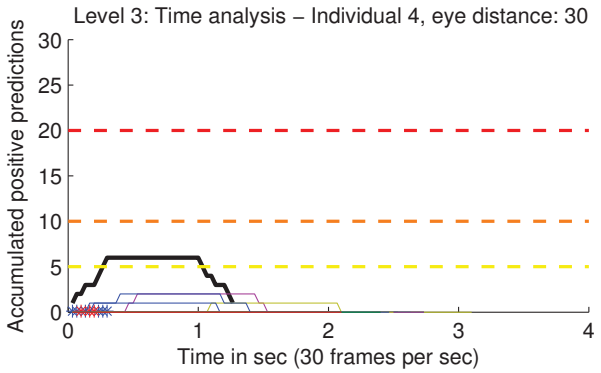

Fig. 7: Accumulated detections for 30 pixels between eyes.

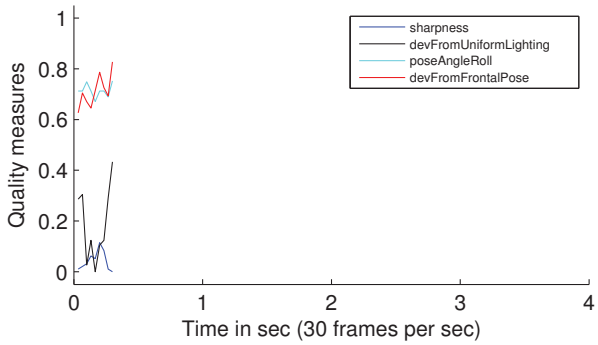

Fig. 8: Variations of quality measures.

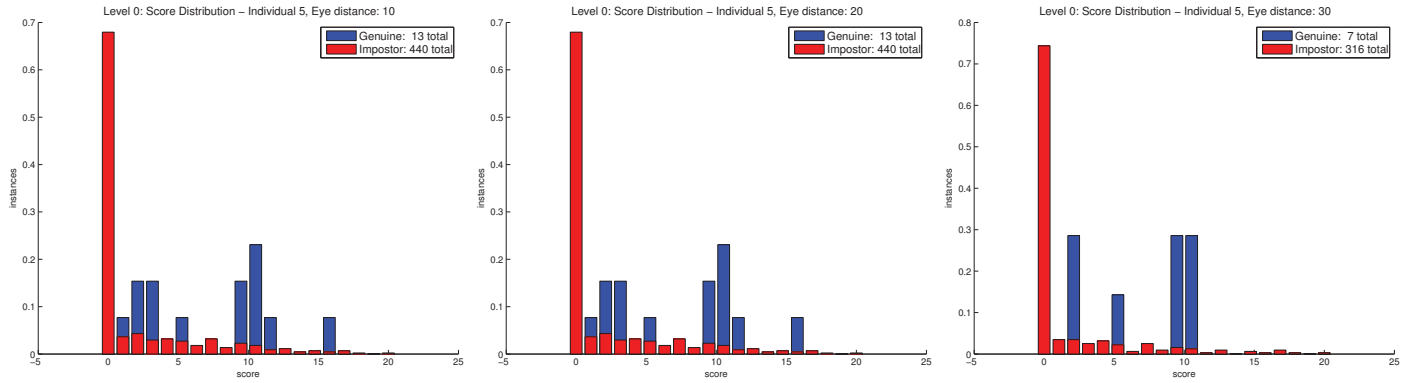

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

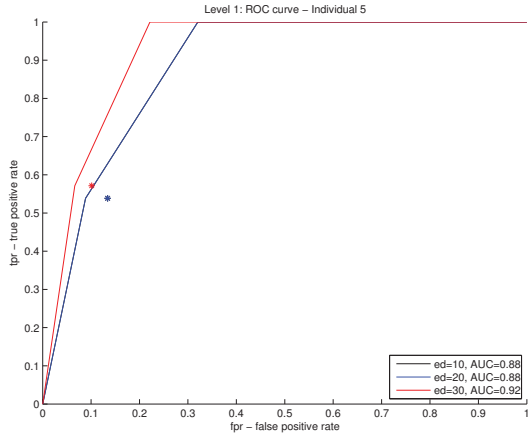

Fig. 2: ROC curve.

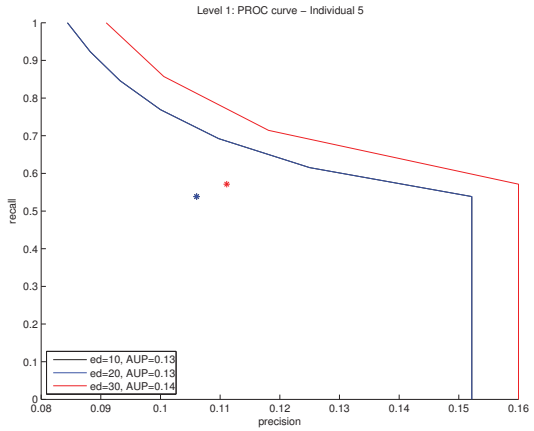

Fig. 3: PROC curve.

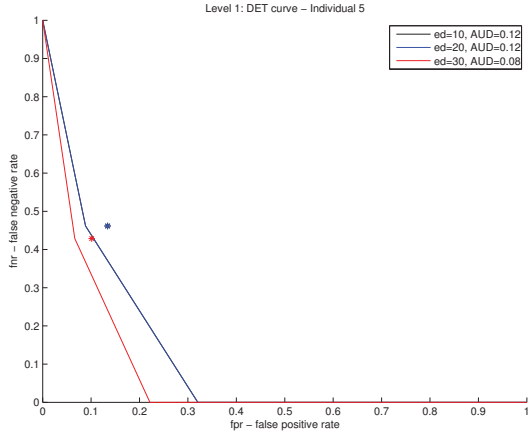

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	13	13	7
Impostor faces (total)	440	440	316
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	7.9976	7.9976	7.2841
False positive rates	13.41%	13.41%	10.13%
True positive rates	53.85%	53.85%	57.14%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	3.12%	21.43%	0.00%	0.00%	**53.85%**	5.26%	13.64%	0.00%	0.00%	0.00%	8.33%	16.67%	39.13%	34.15%	0.00%
20 px.	3.12%	21.43%	0.00%	0.00%	**53.85%**	5.26%	13.64%	0.00%	0.00%	0.00%	8.33%	16.67%	39.13%	34.15%	0.00%
30 px.	0.00%	21.43%	0.00%	0.00%	**57.14%**	5.88%	5.88%	0.00%	0.00%	0.00%	11.76%	22.22%	40.00%	30.43%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	13.33%	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	9.09%	8.82%	8.00%	0.00%	20.00%	17.39%	10.00%	22.22%
20 px.	13.33%	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	9.09%	8.82%	8.00%	0.00%	20.00%	17.39%	10.00%	22.22%
30 px.	8.33%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	9.09%	8.70%	0.00%	0.00%	20.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 5. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

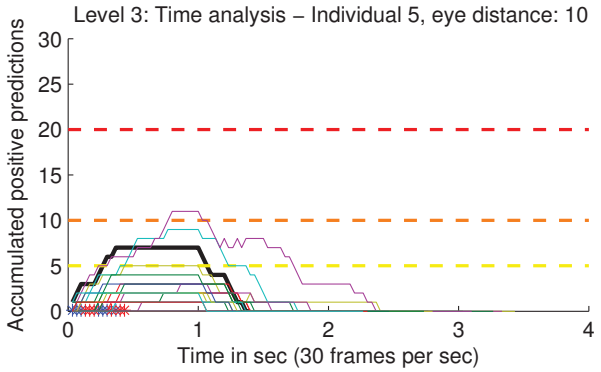

Fig. 5: Accumulated detections for 10 pixels between eyes.

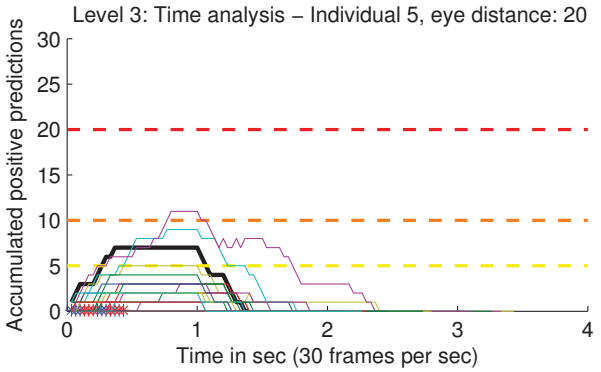

Fig. 6: Accumulated detections for 20 pixels between eyes.

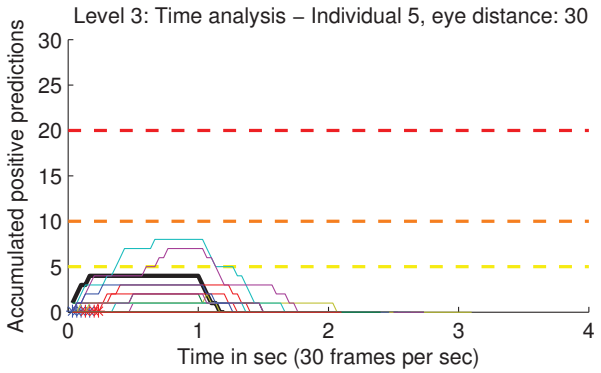

Fig. 7: Accumulated detections for 30 pixels between eyes.

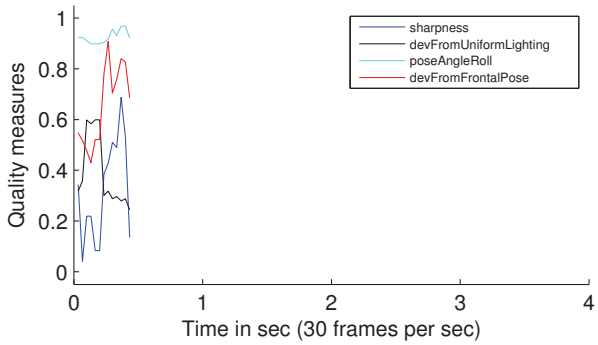

Fig. 8: Variations of quality measures.

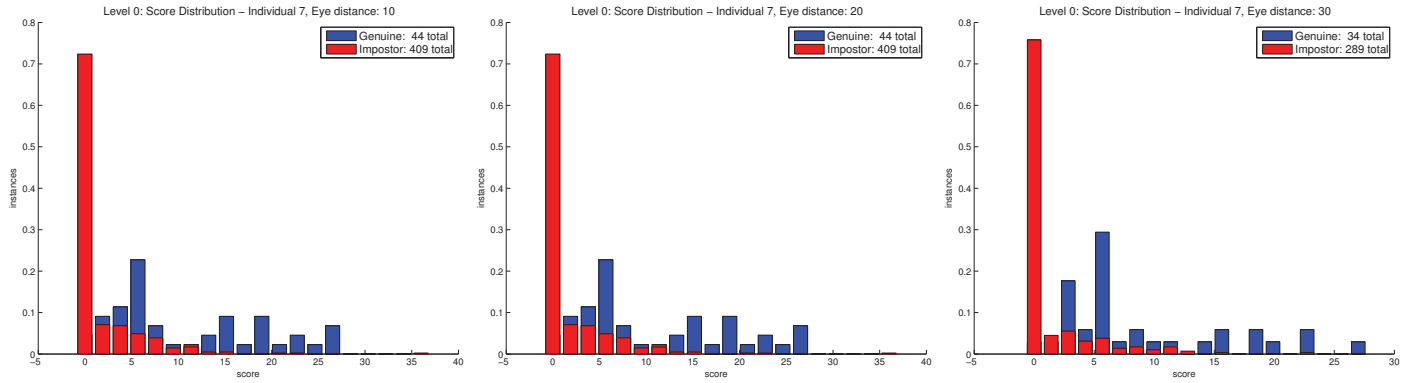

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

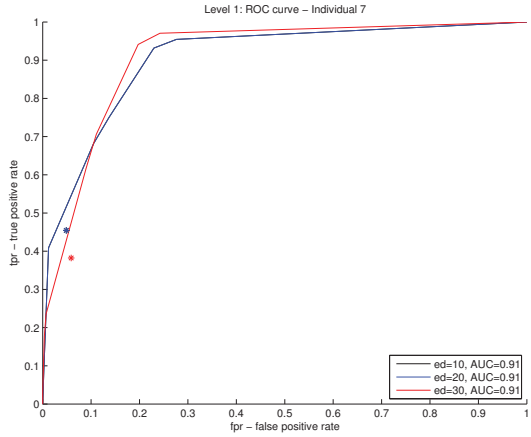

Fig. 2: ROC curve.

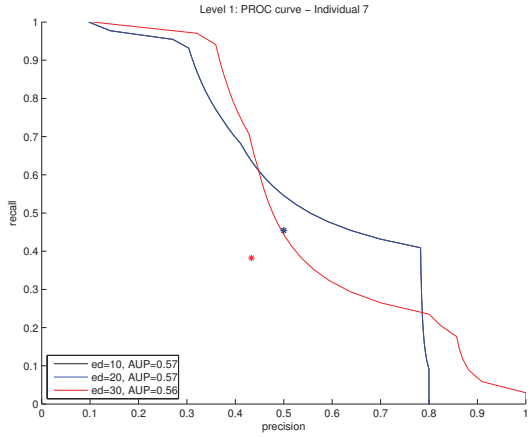

Fig. 3: PROC curve.

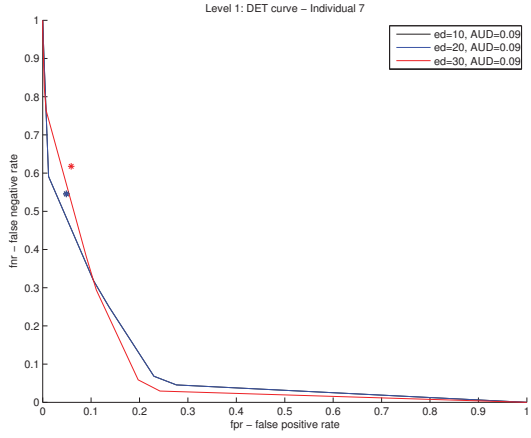

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	44	44	34
Impostor faces (total)	409	409	289
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	9.2545	9.2545	8.4346
False positive rates	4.89%	4.89%	5.88%
True positive rates	45.45%	45.45%	38.24%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	9.38%	0.00%	0.00%	0.00%	0.00%	5.26%	**45.45%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.20%	0.00%
20 px.	9.38%	0.00%	0.00%	0.00%	0.00%	5.26%	**45.45%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	12.20%	0.00%
30 px.	11.76%	0.00%	0.00%	0.00%	0.00%	5.88%	**38.24%**	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	21.74%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	13.33%	0.00%	0.00%	10.26%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%	0.00%	6.67%	4.35%	20.00%	0.00%
20 px.	13.33%	0.00%	0.00%	10.26%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%	0.00%	6.67%	4.35%	20.00%	0.00%
30 px.	16.67%	0.00%	0.00%	13.04%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	13.33%	0.00%	50.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 7. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

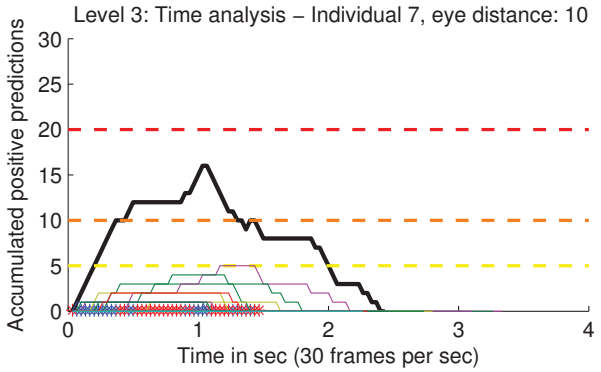

Fig. 5: Accumulated detections for 10 pixels between eyes.

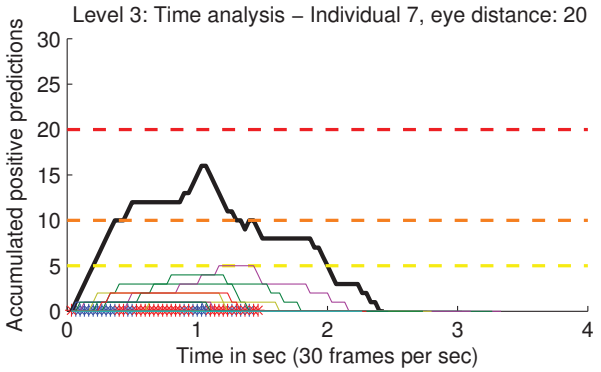

Fig. 6: Accumulated detections for 20 pixels between eyes.

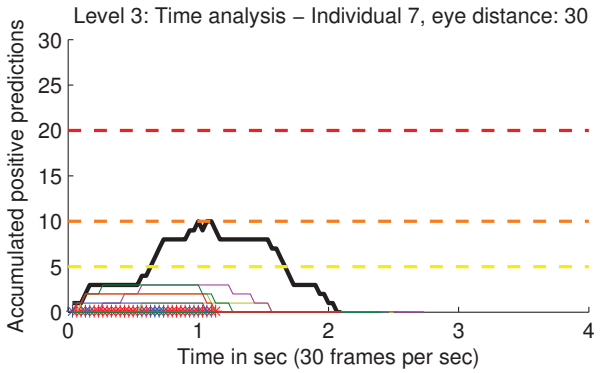

Fig. 7: Accumulated detections for 30 pixels between eyes.

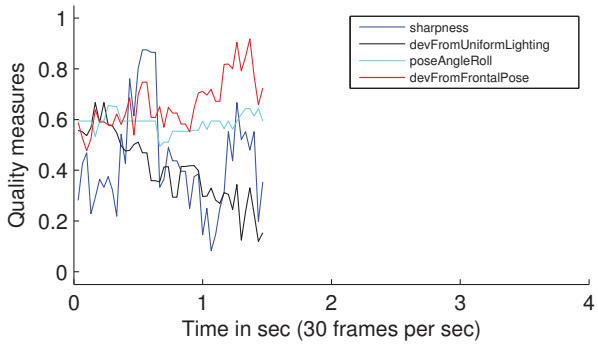

Fig. 8: Variations of quality measures.

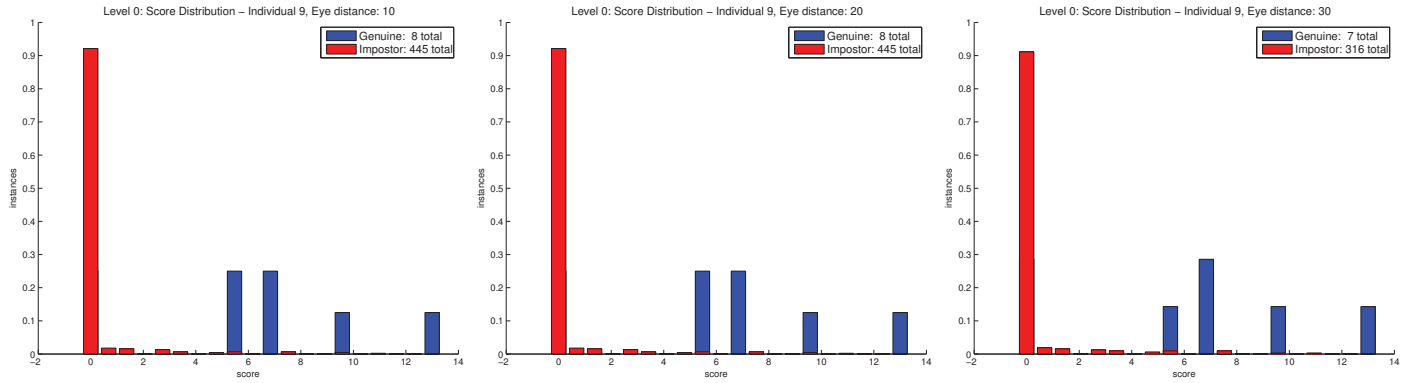

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

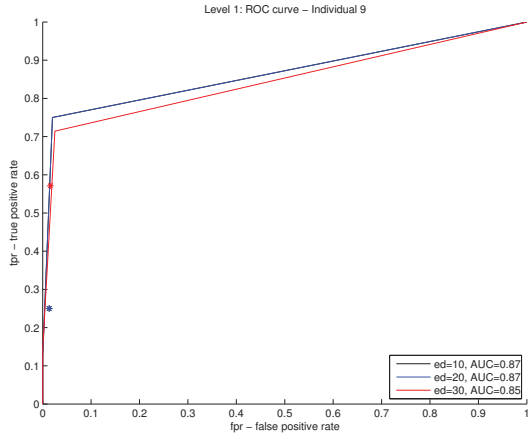

Fig. 2: ROC curve.

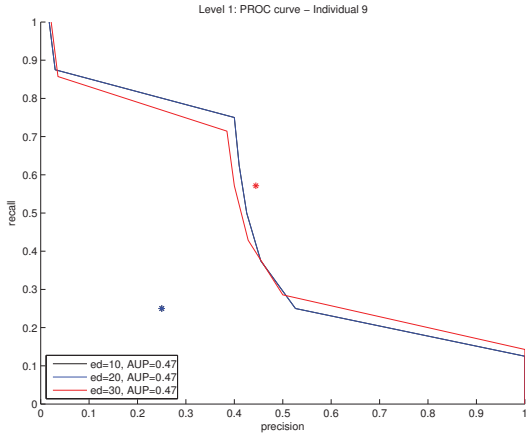

Fig. 3: PROC curve.

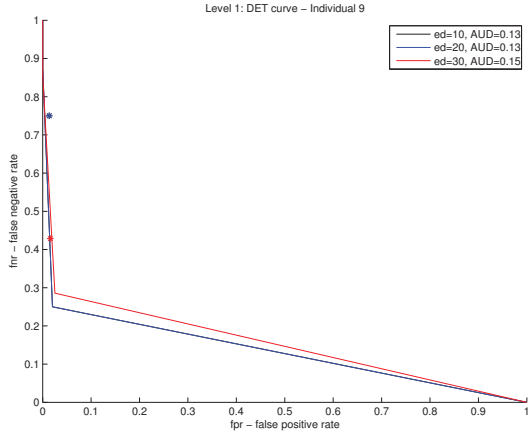

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	8	8	7
Impostor faces (total)	445	445	316
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	7.7002	7.7002	6.7892
False positive rates	1.35%	1.35%	1.58%
True positive rates	25.00%	25.00%	57.14%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	3.12%	0.00%	0.00%	0.00%	0.00%	5.26%	0.00%	0.00%	**25.00**%	0.00%	4.17%	0.00%	0.00%	4.88%	0.00%
20 px.	3.12%	0.00%	0.00%	0.00%	0.00%	5.26%	0.00%	0.00%	**25.00**%	0.00%	4.17%	0.00%	0.00%	4.88%	0.00%
30 px.	5.88%	0.00%	0.00%	0.00%	0.00%	5.88%	0.00%	0.00%	**57.14**%	0.00%	0.00%	0.00%	0.00%	8.70%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 9. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

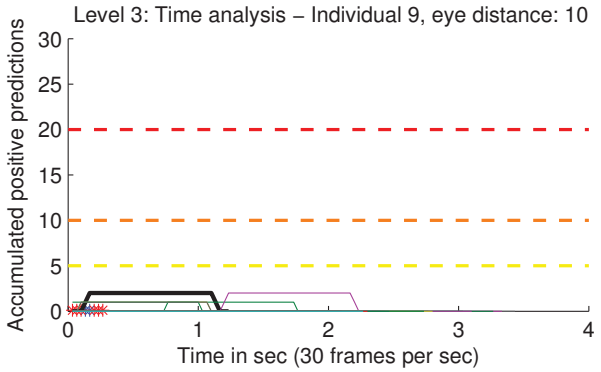

Fig. 5: Accumulated detections for 10 pixels between eyes.

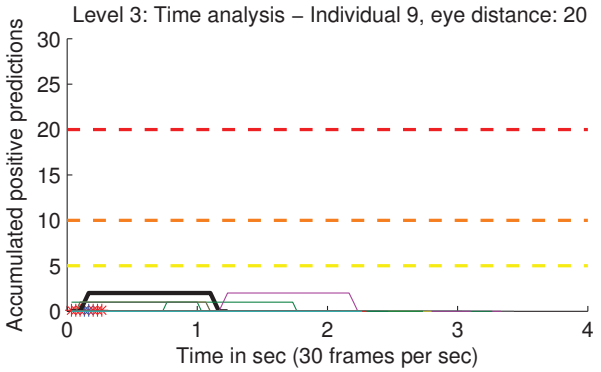

Fig. 6: Accumulated detections for 20 pixels between eyes.

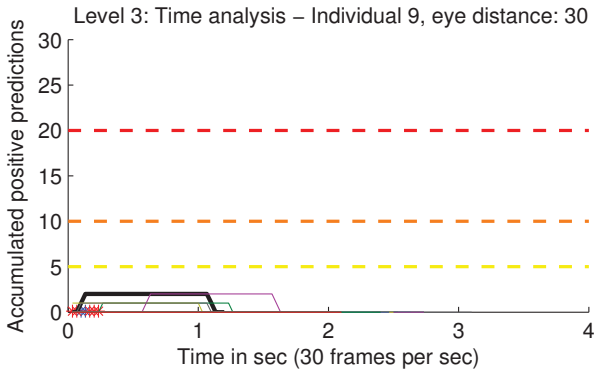

Fig. 7: Accumulated detections for 30 pixels between eyes.

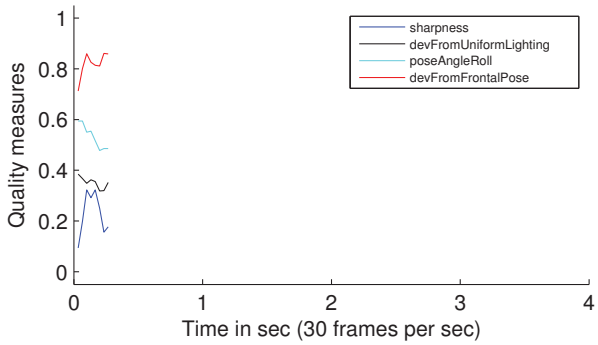

Fig. 8: Variations of quality measures.

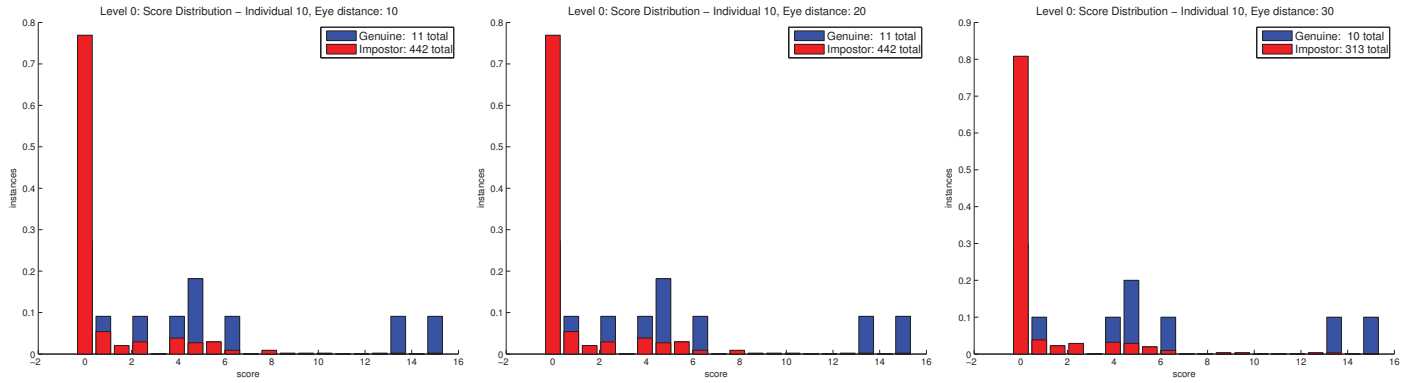

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

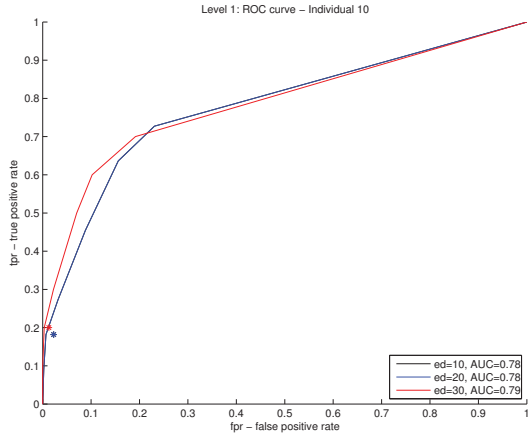

Fig. 2: ROC curve.

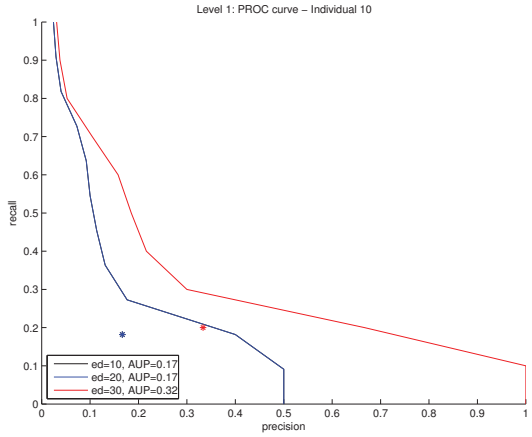

Fig. 3: PROC curve.

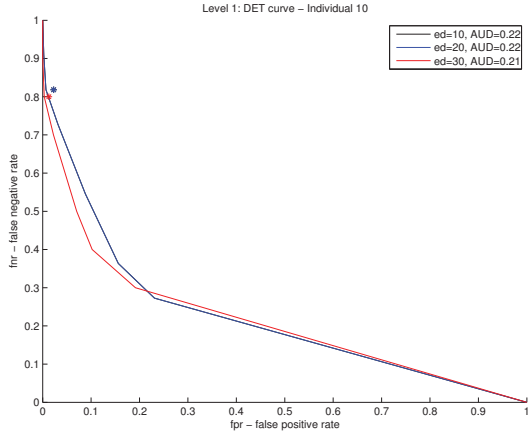

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	11	11	10
Impostor faces (total)	442	442	313
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	7.1200	7.1200	7.2846
False positive rates	2.26%	2.26%	1.28%
True positive rates	18.18%	18.18%	20.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	6.25%	0.00%	0.00%	11.11%	0.00%	0.00%	9.09%	0.00%	0.00%	**18.18%**	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	6.25%	0.00%	0.00%	11.11%	0.00%	0.00%	9.09%	0.00%	0.00%	**18.18%**	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	5.88%	0.00%	0.00%	11.11%	0.00%	0.00%	0.00%	0.00%	0.00%	**20.00%**	0.00%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.94%	4.00%	0.00%	0.00%	0.00%	10.00%	0.00%
20 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	2.94%	4.00%	0.00%	0.00%	0.00%	10.00%	0.00%
30 px.	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.35%	6.25%	0.00%	0.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 10. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

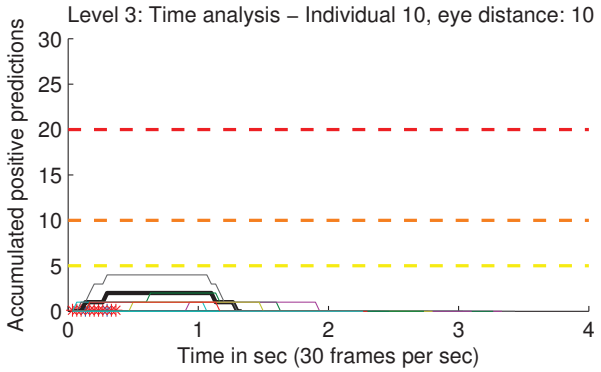

Fig. 5: Accumulated detections for 10 pixels between eyes.

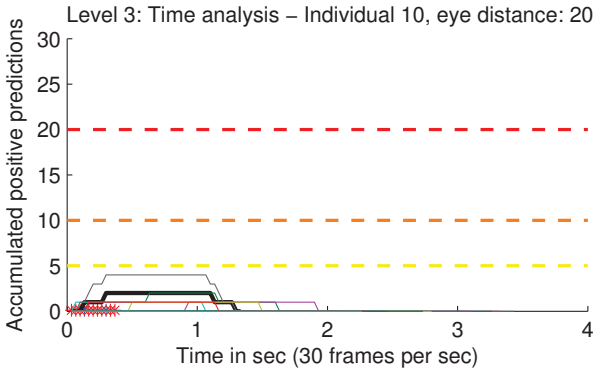

Fig. 6: Accumulated detections for 20 pixels between eyes.

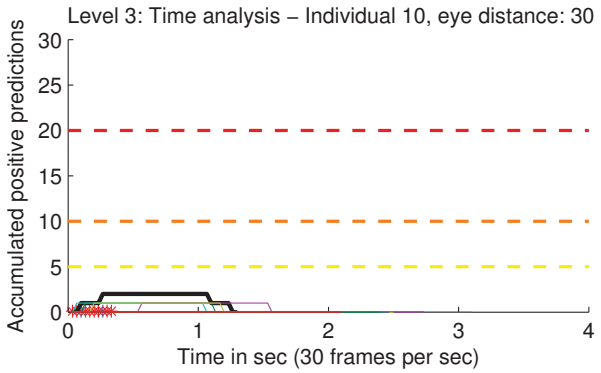

Fig. 7: Accumulated detections for 30 pixels between eyes.

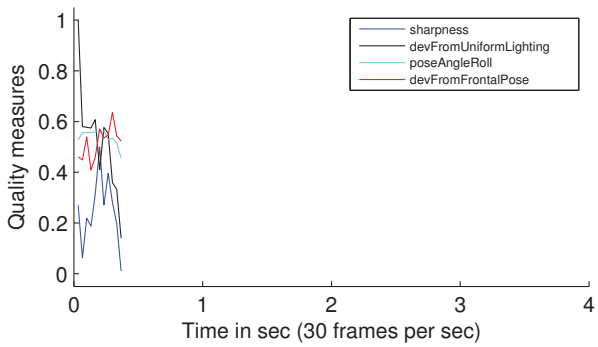

Fig. 8: Variations of quality measures.

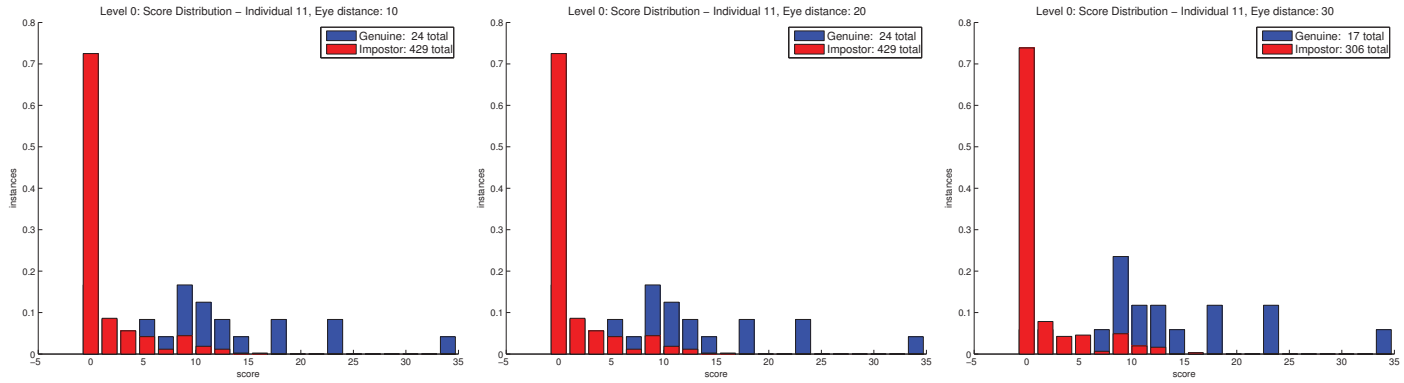

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

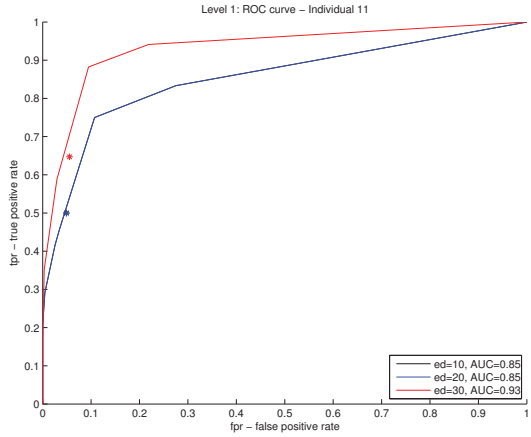

Fig. 2: ROC curve.

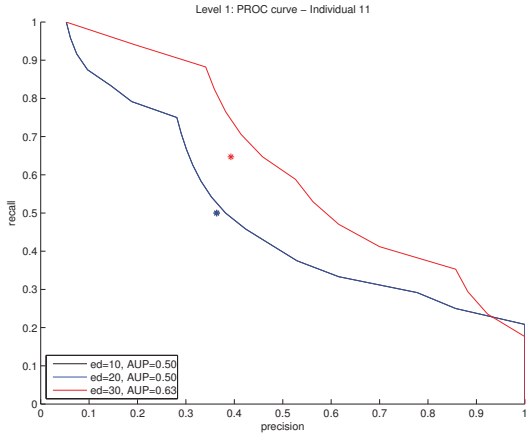

Fig. 3: PROC curve.

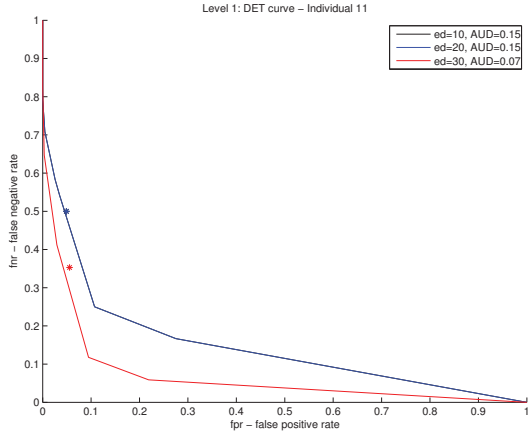

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	24	24	17
Impostor faces (total)	429	429	306
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	9.0293	9.0293	9.5547
False positive rates	4.90%	4.90%	5.56%
True positive rates	50.00%	50.00%	64.71%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	6.25%	28.57%	0.00%	0.00%	7.69%	5.26%	6.82%	0.00%	0.00%	9.09%	**50.00%**	0.00%	0.00%	0.00%	0.00%
20 px.	6.25%	28.57%	0.00%	0.00%	7.69%	5.26%	6.82%	0.00%	0.00%	9.09%	**50.00%**	0.00%	0.00%	0.00%	0.00%
30 px.	11.76%	28.57%	0.00%	0.00%	0.00%	5.88%	5.88%	0.00%	0.00%	10.00%	**64.71%**	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	13.33%	0.00%	0.00%	2.56%	0.00%	0.00%	0.00%	0.00%	0.00%	8.00%	0.00%	0.00%	13.04%	10.00%	0.00%
20 px.	13.33%	0.00%	0.00%	2.56%	0.00%	0.00%	0.00%	0.00%	0.00%	8.00%	0.00%	0.00%	13.04%	10.00%	0.00%
30 px.	16.67%	0.00%	0.00%	4.35%	0.00%	0.00%	0.00%	0.00%	0.00%	6.25%	0.00%	0.00%	23.08%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 11. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

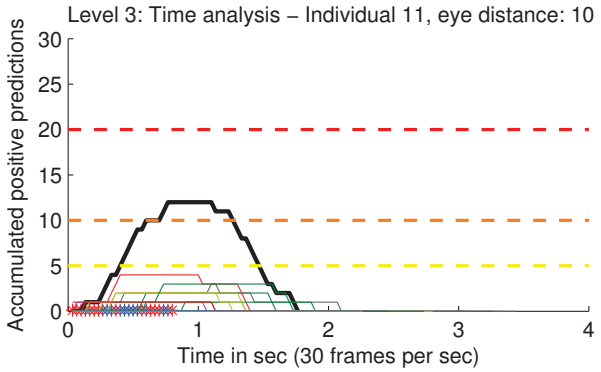

Fig. 5: Accumulated detections for 10 pixels between eyes.

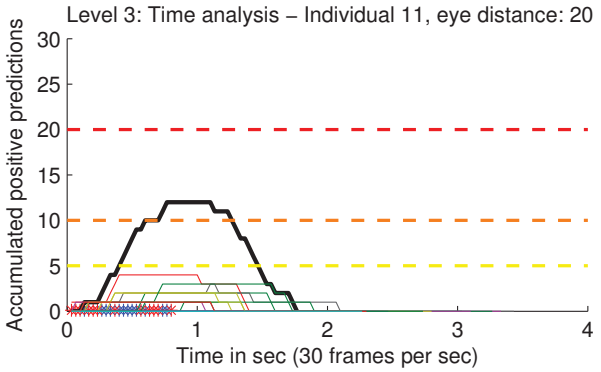

Fig. 6: Accumulated detections for 20 pixels between eyes.

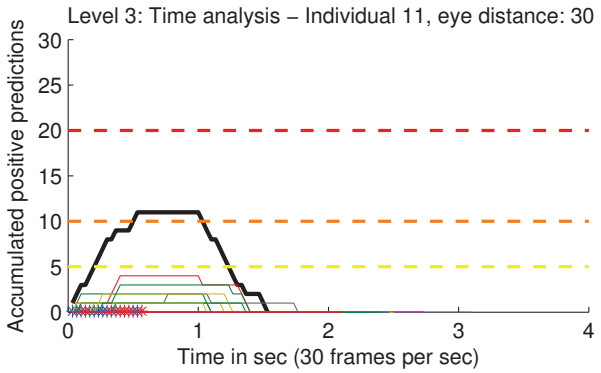

Fig. 7: Accumulated detections for 30 pixels between eyes.

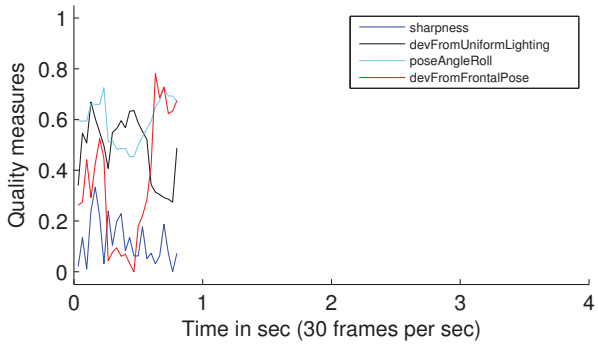

Fig. 8: Variations of quality measures.

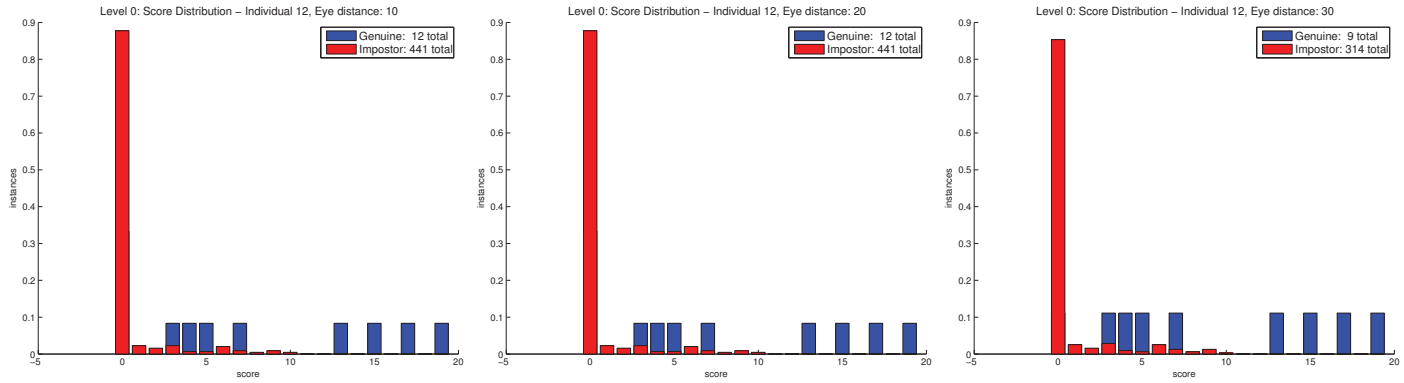

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

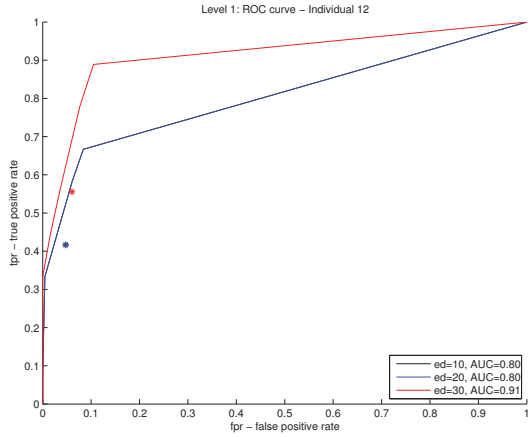

Fig. 2: ROC curve.

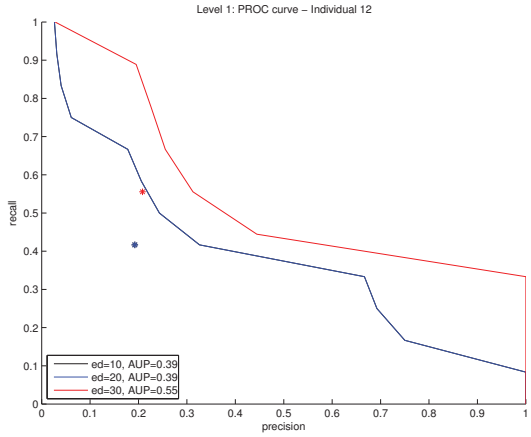

Fig. 3: PROC curve.

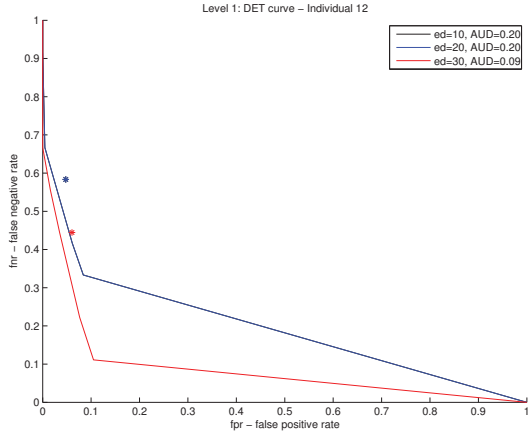

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	12	12	9
Impostor faces (total)	441	441	314
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	5.6674	5.6674	5.8000
False positive rates	4.76%	4.76%	6.05%
True positive rates	41.67%	41.67%	55.56%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	21.88%	0.00%	0.00%	33.33%	0.00%	5.26%	0.00%	0.00%	0.00%	0.00%	0.00%	**41.67%**	0.00%	2.44%	0.00%
20 px.	21.88%	0.00%	0.00%	33.33%	0.00%	5.26%	0.00%	0.00%	0.00%	0.00%	0.00%	**41.67%**	0.00%	2.44%	0.00%
30 px.	29.41%	0.00%	0.00%	33.33%	0.00%	5.88%	0.00%	0.00%	0.00%	0.00%	0.00%	**55.56%**	0.00%	4.35%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	20.59%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%
20 px.	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	20.59%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%
30 px.	0.00%	0.00%	7.69%	0.00%	0.00%	0.00%	0.00%	0.00%	30.43%	0.00%	25.00%	0.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 12. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

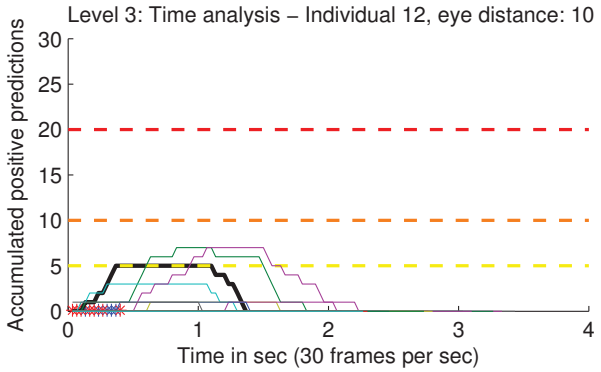

Fig. 5: Accumulated detections for 10 pixels between eyes.

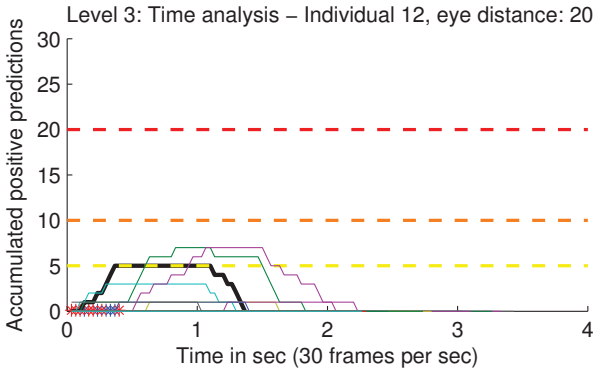

Fig. 6: Accumulated detections for 20 pixels between eyes.

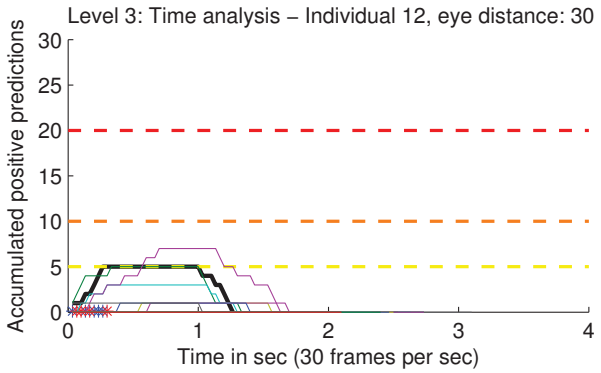

Fig. 7: Accumulated detections for 30 pixels between eyes.

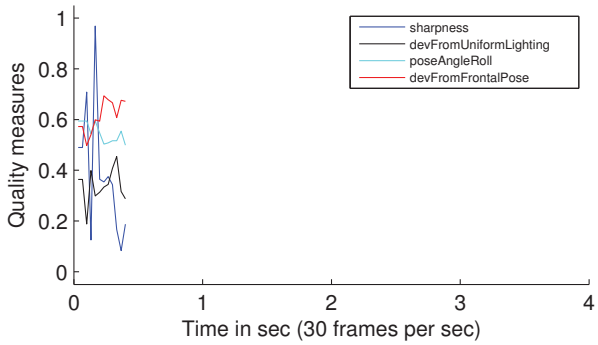

Fig. 8: Variations of quality measures.

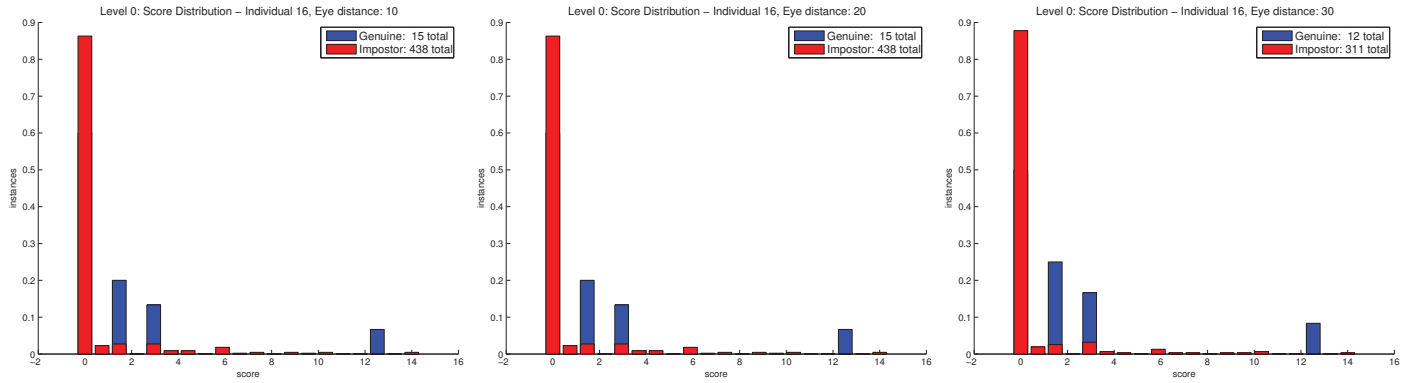

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

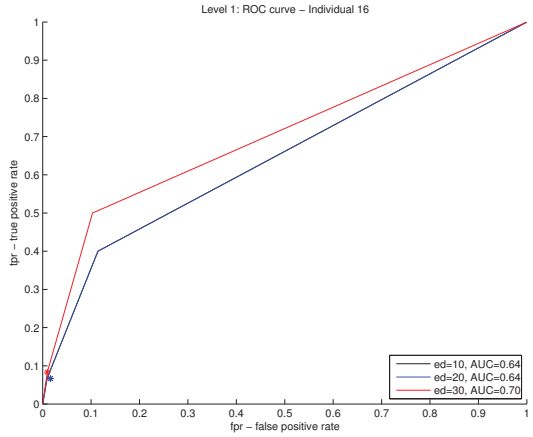

Fig. 2: ROC curve.

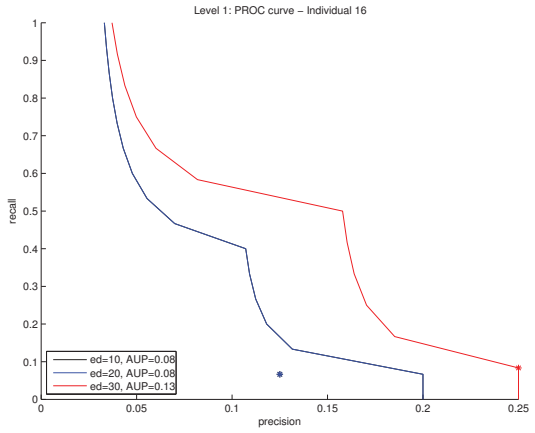

Fig. 3: PROC curve.

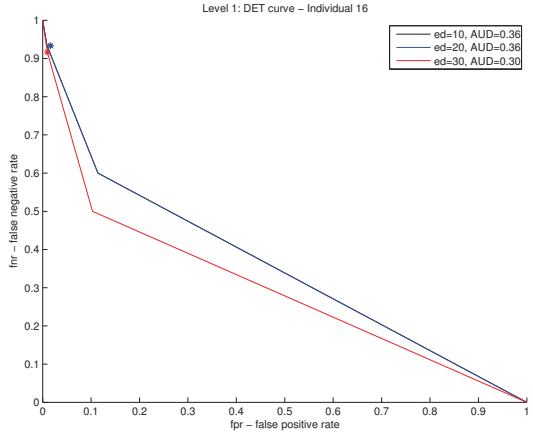

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	15	15	12
Impostor faces (total)	438	438	311
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	8.7424	8.7424	10.7727
False positive rates	1.60%	1.60%	0.96%
True positive rates	6.67%	6.67%	8.33%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	6.25%	7.14%	0.00%	0.00%	0.00%	0.00%	2.27%	0.00%	0.00%	0.00%	4.17%	0.00%	0.00%	2.44%	0.00%
20 px.	6.25%	7.14%	0.00%	0.00%	0.00%	0.00%	2.27%	0.00%	0.00%	0.00%	4.17%	0.00%	0.00%	2.44%	0.00%
30 px.	5.88%	0.00%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	4.35%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	**6.67**%	0.00%	0.00%	2.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
20 px.	**6.67**%	0.00%	0.00%	2.56%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
30 px.	**8.33**%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 16. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

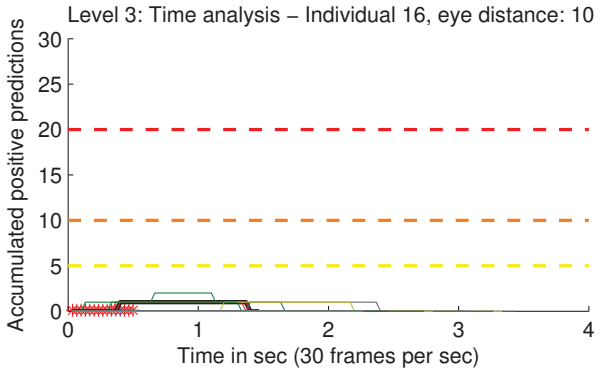

Fig. 5: Accumulated detections for 10 pixels between eyes.

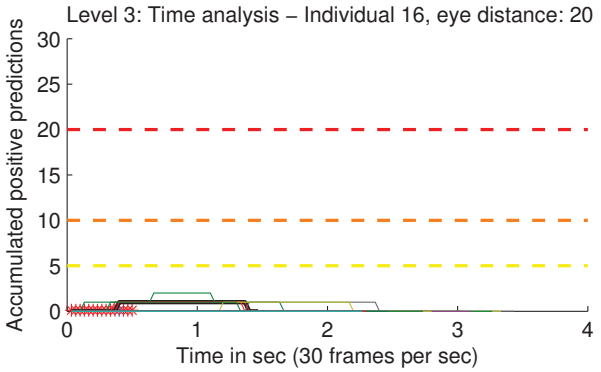

Fig. 6: Accumulated detections for 20 pixels between eyes.

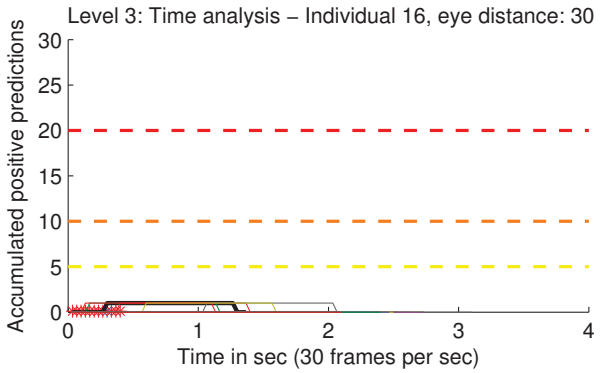

Fig. 7: Accumulated detections for 30 pixels between eyes.

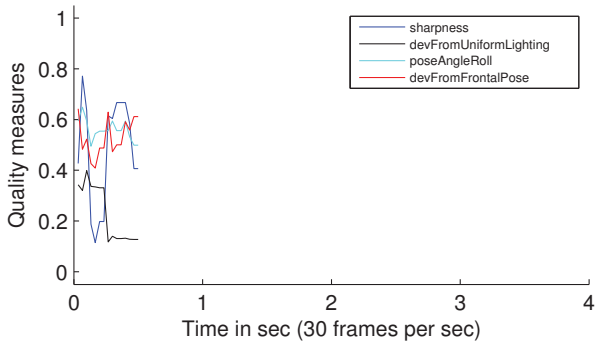

Fig. 8: Variations of quality measures.

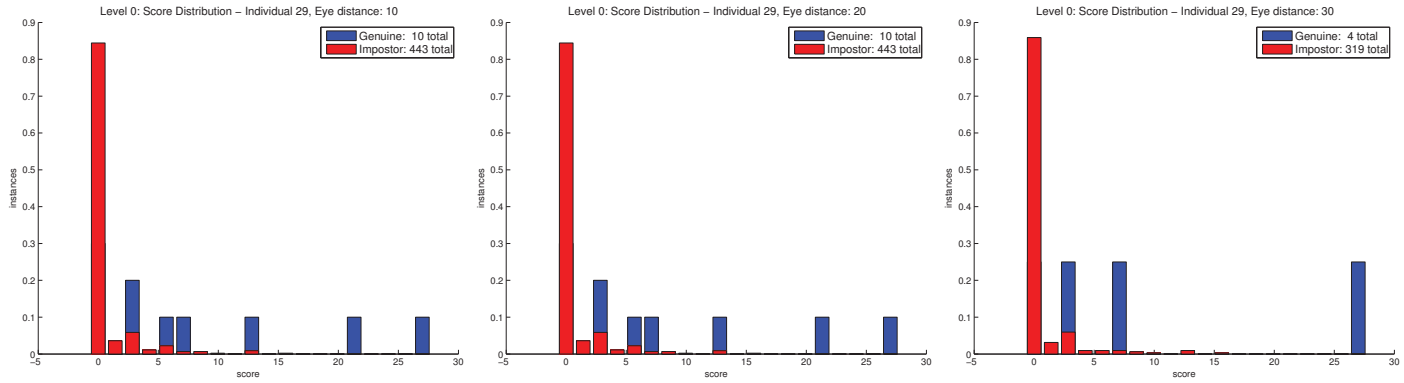

Fig. 1: Level 0 – Class scores distributions.

Level 1 Analysis

The figures bellow detail several performance curves, and the stars indicate the selected operation point for a target $fpr = 5\%$ (for each ed distance between eyes). The table summarizes the number of genuine and impostor samples, as well as face detection related metrics.

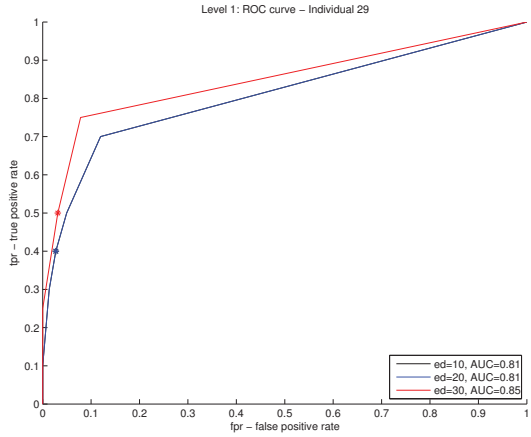

Fig. 2: ROC curve.

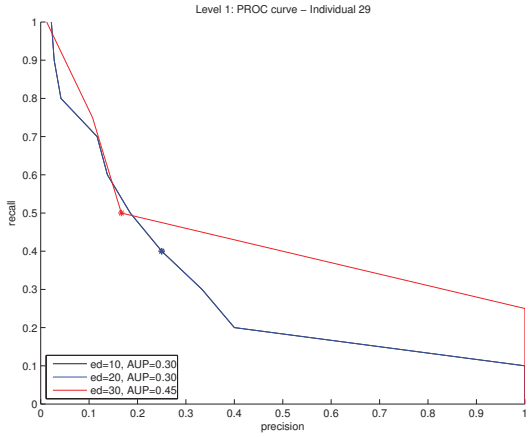

Fig. 3: PROC curve.

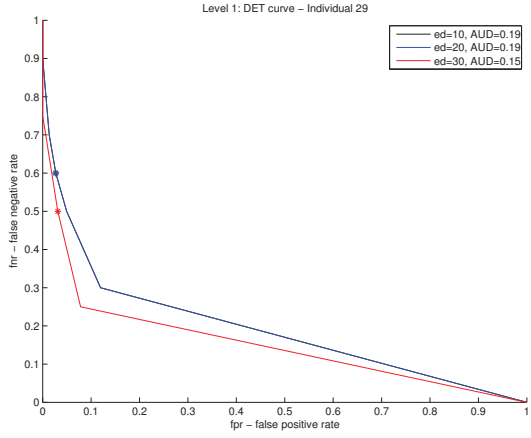

Fig. 4: DET curve.

Measure	Eyes distance (pixels)		
	10	**20**	**30**
Genuine faces (total)	10	10	4
Impostor faces (total)	443	443	319
Detection Level			
Falsely detected faces	0.22%	0.22%	0.31%
Failure to acquire rate	74.48%	74.48%	81.86%
Matching Level			
Low quality faces	0.00%	0.00%	0.00%
Operating points	7.5750	7.5750	7.0853
False positive rates	2.71%	2.71%	3.13%
True positive rates	40.00%	40.00%	50.00%

Tab. 1: Test set results for $fpr = 5\%$.

Level 2 Analysis

Distance	Ind. 1	Ind. 2	Ind. 3	Ind. 4	Ind. 5	Ind. 6	Ind. 7	Ind. 8	Ind. 9	Ind. 10	Ind. 11	Ind. 12	Ind. 13	Ind. 14	Ind. 15
10 px.	6.25%	7.14%	0.00%	0.00%	0.00%	5.26%	0.00%	0.00%	0.00%	0.00%	8.33%	0.00%	0.00%	2.44%	0.00%
20 px.	6.25%	7.14%	0.00%	0.00%	0.00%	5.26%	0.00%	0.00%	0.00%	0.00%	8.33%	0.00%	0.00%	2.44%	0.00%
30 px.	11.76%	7.14%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%	11.76%	0.00%	0.00%	0.00%	0.00%

(a)

Distance	Ind. 16	Ind. 17	Ind. 18	Ind. 19	Ind. 20	Ind. 21	Ind. 22	Ind. 23	Ind. 24	Ind. 25	Ind. 26	Ind. 27	Ind. 28	Ind. 29	Ind. 30
10 px.	13.33%	0.00%	7.69%	2.56%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%	0.00%	0.00%	0.00%	**40.00%**	0.00%
20 px.	13.33%	0.00%	7.69%	2.56%	0.00%	0.00%	0.00%	0.00%	2.94%	0.00%	0.00%	0.00%	0.00%	**40.00%**	0.00%
30 px.	16.67%	0.00%	7.69%	4.35%	0.00%	0.00%	0.00%	0.00%	4.35%	0.00%	0.00%	0.00%	0.00%	**50.00%**	0.00%

(b)

Tab. 2: Dodington's zoo based analysis for the detection module associated to individual 29. Columns details the individuals in the data set, while lines detail their detection by the module for each value of distance between the eyes. Colors are as follows: green for sheep like individuals (easy to predict), yellow for goat like individuals (difficult to predict), blue for lamb like individuals (can be impersonated by someone else) and red for wolf like individuals (who can impersonate another user).

Level 3 Analysis

Each of the below figures details the performance of systems by accumulating positive predictions over a time-window on the video stream for different distances between the eyes. The tracker is used to separate faces of different persons, and accumulate their predictions. Matching thresholds are set to provide a 5% false positive rate, and positive individual decision takes place after accumulating 20 detections in a 30 frames window (1 sec). Red stars indicate faces that have not been correctly matched to the target individual, while blue stars indicate that the individual captured in the video has been successfully matched to the target individual.

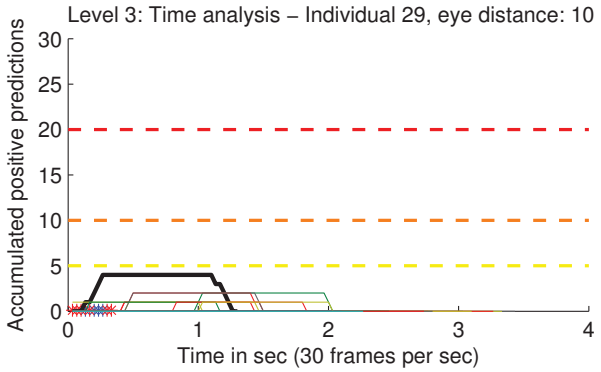

Fig. 5: Accumulated detections for 10 pixels between eyes.

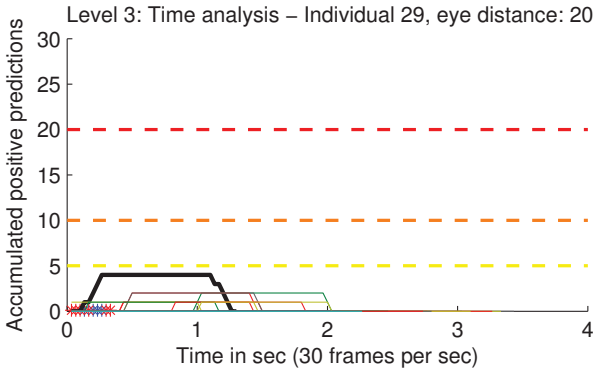

Fig. 6: Accumulated detections for 20 pixels between eyes.

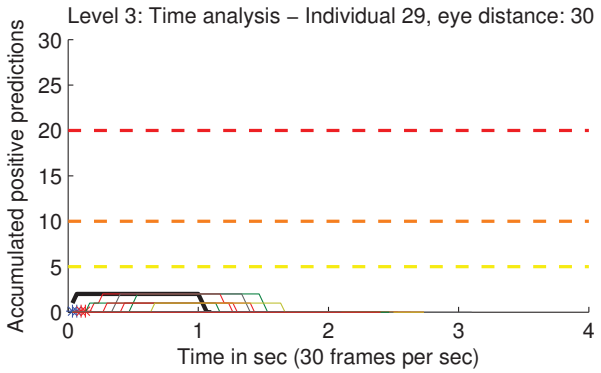

Fig. 7: Accumulated detections for 30 pixels between eyes.

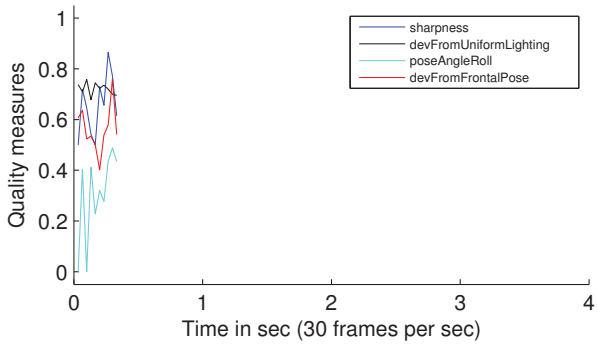

Fig. 8: Variations of quality measures.